Synchronization Problems in Computer Vision

Federica Arrigoni[2] and Andrea Fusiello[1]

[1] DPIA - Università di Udine Via Delle Scienze, 208 - Udine, Italy
andrea.fusiello@uniud.it
[2] Czech Technical University in Prague
arrigoni.federica@spes.uniud.it

Abstract. The goal of synchronization is to infer the unknown states of a network of nodes, where only the ratio (or difference) between pairs of states can be measured. Typically, states are represented by elements of a group, such as the Symmetric Group or the Special Euclidean Group. The former can represent local labels of a set of features, which refer to the multi-view matching application, whereas the latter can represent camera reference frames, in which case we are in the context of structure from motion, or local coordinates where 3D points are represented, in which case we are dealing with multiple point-set registration. A related problem is that of bearing-based network localization where each node is located at a fixed (unknown) position in 3-space and pairs of nodes can measure the direction of the line joining their locations.

This paper is concerned with the *synchronization* problem, which can be stated as follows: given a network of nodes, where each node is characterized by an unknown state and pairs of nodes can measure the ratio (or difference) between their states, the goal is to estimate the unknown states from the pairwise measures. The problem can be modeled as a graph where nodes correspond to the unknown states and edges encode the pairwise measures, and it is well-posed only if such a graph is connected.

As an example, consider the graph in Figure 1, where nodes and edges are labelled with integer numbers: the task is to recover the unknown numbers in the nodes by measuring their differences (on the edges). Two things can be immediately observed: a solution exists only if the sum of the differences along any cycle is zero, and, when it exists, the solution is not unique, for adding a constant to the nodes does not change the differences.

Mathematically, states are represented by elements of a group Σ: $\Sigma = \mathbb{R}$ yields *time synchronization* [1, 2], from which the term *synchronization* originates, where all the nodes in a network are synchronized to a common clock. $\Sigma = SO(d)$ corresponds to *rotation synchronization* (also known as rotation averaging) [3–13] and $\Sigma = SE(d)$ results in *rigid-motion synchronization* (also known as motion averaging or pose graph optimization) [14–22], which find application in structure from motion, registration of 3D point sets and simultaneous localization and mapping.

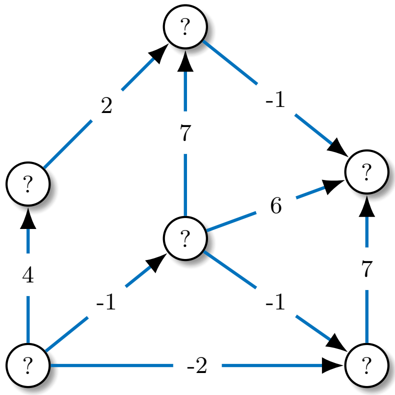

Fig. 1. synchronization over $(\mathbb{Z}, +)$.

Finally, $\Sigma = \mathcal{S}_d$ and $\Sigma = \mathcal{I}_d$ give rise to *permutation synchronization* [23–25] and *partial permutation synchronization* [26], respectively, which are related to multi-view matching.

In [27] we set forth a theoretical unified framework where several synchronization problems are seen as instances of a more abstract principle. Thanks to the formalism of synchronization, several Computer Vision problems can be addressed without relying on features or points, since the problem is formulated in frame space, or, more abstractly, in a group. In this paper we will focus on permutation synchronization and motion synchronization.

In the case of *permutation synchronization*, each state is an unknown reordering of d objects, which is represented as a $d \times d$ permutation matrix, namely an element of the Symmetric Group $Sym(d)$. It is assumed that pairs of nodes can match these objects, establishing which objects are the same in the two nodes, despite the different naming, and the goal is to infer a global labeling of the objects, such that the same object receives the same label in all the nodes (see Fig. 2).

Permutation synchronization finds application in *multi-view matching*, where nodes are images and objects are features. In practice, not all the features are visible in all the images, so matches are modeled as *partial* permutations, which form the so-called Symmetric Inverse Semigroup $ISym(d)$. In [28, 26] we develop a novel solution to partial permutation synchronization based on a spectral decomposition, which successfully handles missing correspondences.

In the case of *rigid-motion synchronization*, each state is the angular attitude and position of a d-dimensional reference frame, which is referred to as *motion* in Computer Vision, *orientation* in Photogrammetry, or *pose* in Robotics, and it is described by a direct isometry, which is an element of the Special Euclidean Group $SE(d)$.

The goal is to recover the location and attitude of a set of reference frames organized in a network, where the links of this network are relative transforma-

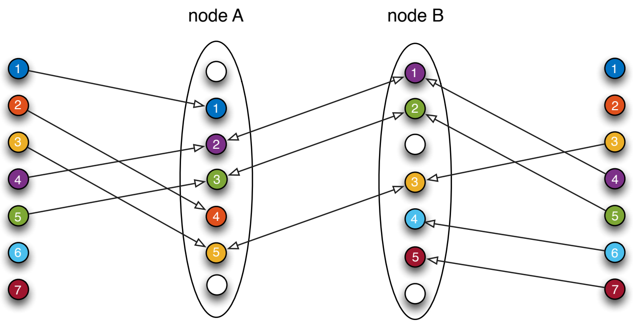

Fig. 2. In the center, two nodes with partial visibility match their three common objects. At the extrema the ground truth ordering of the objects. Each node sees some of the objects (white circles are missing objects) and puts them in a different order, i.e., it gives them different numeric labels.

tions of one frame with respect to the others. If we restrict the attention to the angular attitude (leaving out the position) then we get rotation synchronization. Similarly, if position only is considered, it results in translation synchronization.

Such local frames can be camera reference frames, in which case we are in the context of *structure-from-motion* (see Fig. 3), or local coordinates where 3D points are represented, in which case we are dealing with *multiple point-set registration* (see Fig. 4). In the first case, the goal is to recover both the 3D structure of the scene and camera motion from multiple images, whereas in the second case the goal is to find the rigid transformations that bring multiple 3D point sets into alignment.

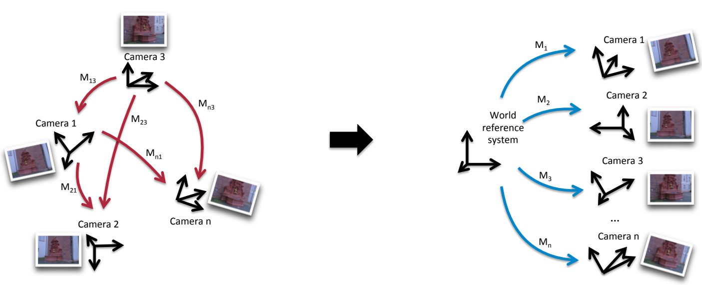

Fig. 3. Structure from motion with motion synchronization.

In [29, 20] we express synchronization over $SO(d)$ and $SE(d)$ in terms of *low-rank and sparse* (LRS) matrix decomposition, that is the problem of recovering a low-rank matrix starting from an incomplete subset of its entries, possibly corrupted by noise and outliers. As far as the Special Euclidean Group is concerned,

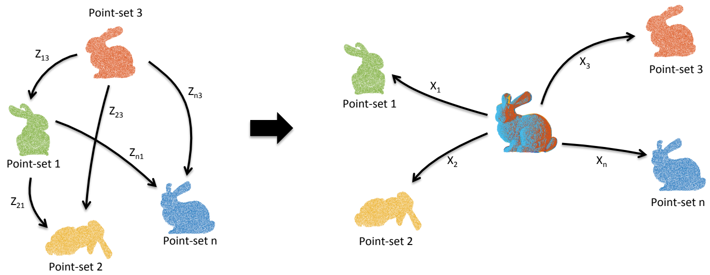

Fig. 4. Multiple point-set registration with motion synchronization

we tackle the problem of multiple point-set registration, whereas in the case of the Special Orthogonal Group we concentrate on structure from motion. In [19] we also propose a closed-form approach to synchronization over $SE(d)$ based on a spectral decomposition, which is applied to multiple point-set registration, where robustness to outliers is be gained via iteratively re-weighted least squares (IRLS).

This method can be viewed as the extension to $SE(d)$ of the spectral solution developed in [5, 30, 31] for $SO(d)$.

The motion task of the structure-from-motion problem cannot be straightforwardly solved as a synchronization over $SE(3)$, due to the depth-speed ambiguity: only the *directions* of the relative displacements between camera pairs can be measured, but the magnitude is unknown.

A possibility consists in breaking the motion estimation process in two stages, namely a rotation synchronization to obtain the angular attitudes of the cameras, followed by the recovery of camera positions from pairwise directions, which is an instance of *bearing-based network localization* in 3-space.

This workflow is exploited in several structure from motion systems, such as [32–36].

Alternatively, the translation magnitudes (referred to as the *epipolar scales*) can be explicitly computed, as we propose in [37], via a two-stage method: first, a *cycle basis* is computed; then, all the epipolar scales are recovered simultaneously by solving a homogeneous linear system .

This allows either to address the synchronization problem over $SE(3)$, using (e.g.) the spectral solution [19], or to perform rotation synchronization followed by translation synchronization. With reference to the latter, we also propose a "divide and conquer" technique for computing the epipolar scales [38].

The two paths are equivalent: in [27] we show that the epipolar scales can be uniquely (up to scale) recovered from pairwise directions if and only if node locations can be uniquely (up to translation and scale) recovered from pairwise directions. Requiring that the graph is connected is not sufficient for unique localizability, but more complicated assumptions are required, which are studied under the name of *parallel rigidity*. Several theoretical results about parallel rigidity are present in the literature, which come from disparate communities, including discrete geometry, computer vision, and robotics.

Exhaustive experiments were performed to evaluate the proposed solutions on both synthetic scenarios and real data in the context of structure from motion, multiple point-set registration and multi-view matching.

In general synchronization translate into efficient closed form solutions, namely spectral decomposition or linear least squares. As a matter of fact, in the experiments reported in the literature [30, 18, 19] they are consistently the fastest method.

For instance, the motion synchronization pipeline in [38], takes 7 seconds for the Madrid Metropolis dataset [35], whose epipolar graph contains about 300 nodes and 65% of missing edges (with respect to the complete graph).

Besides the least-squares solution for translation synchronization, which is statistically optimal, the spectral (or null-space) solution of the other instances of synchronization provides an extrinsic estimate, whose quality is – in general – inferior to those provided by intrinsic methods.

For instance, in [19] the spectral method is compared to [21], which minimizes a geometric error tightly related to the synchronization cost function. The latter returns a more accurate solution, but it requires a significant amount of time. However, the accuracy obtained by the spectral method is high (although not optimal), as demonstrated by experiments performed in [19] in a variety of scenarios. As a consequence, it can be seen as a good and fast initialization for a subsequent local refinement (e.g. bundle adjustment in structure from motion).

The spectral solution can also be made resistant to outliers via Iteratively Reweighted Least Squares (IRLS) [39], as suggested in [19], thanks to the fact that the matrix formulation of synchronization can easily cope with weights on individual edge labels (by means of a *weighted* adjacency matrix).

References

1. Karp, R., Elson, J., Estrin, D., Shenker, S.: Optimal and global time synchronization in sensor nets. Technical report, Center for Embedded Networked Sensing: University of California, Los Angeles (2003)
2. Giridhar, A., Kumar, P.: Distributed clock synchronization over wireless networks: Algorithms and analysis. Proceedings of the IEEE Conference on Decision and Control (2006) 4915–4920
3. Sharp, G.C., Lee, S.W., Wehe, D.K.: Multiview registration of 3D scenes by minimizing error between coordinate frames. In: Proceedings of the European Conference on Computer Vision. (2002) 587–597
4. Martinec, D., Pajdla, T.: Robust rotation and translation estimation in multiview reconstruction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2007)
5. Singer, A.: Angular synchronization by eigenvectors and semidefinite programming. Applied and Computational Harmonic Analysis **30**(1) (2011) 20 – 36
6. Crandall, D., Owens, A., Snavely, N., Huttenlocher, D.P.: Discrete-continuous optimization for large-scale structure from motion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2011) 3001–3008

7. Fredriksson, J., Olsson, C.: Simultaneous multiple rotation averaging using lagrangian duality. In: Proceedings of the Asian Conference on Computer Vision. (2012)
8. Boumal, N., Singer, A., Absil, P.A.: Robust estimation of rotations from relative measurements by maximum likelihood. In: Proceedings of the IEEE International Conference on Robotics and Automation. (2013)
9. Hartley, R., Trumpf, J., Dai, Y., Li, H.: Rotation averaging. International Journal of Computer Vision (2013)
10. Chatterjee, A., Govindu, V.M.: Efficient and robust large-scale rotation averaging. In: Proceedings of the International Conference on Computer Vision. (2013)
11. Wang, L., Singer, A.: Exact and stable recovery of rotations for robust synchronization. Information and Inference: a Journal of the IMA **2**(2) (2013) 145–193
12. Arrigoni, F., Magri, L., Rossi, B., Fragneto, P., Fusiello, A.: Robust absolute rotation estimation via low-rank and sparse matrix decomposition. In: Proceedings of the International Conference on 3D Vision (3DV). (2014) 491–498
13. Tron, R., Zhou, X., Daniilidis, K.: A survey on rotation optimization in structure from motion. In: Computer Vision and Pattern Recognition Workshops (CVPRW). (2016)
14. Fusiello, A., Castellani, U., Ronchetti, L., Murino, V.: Model acquisition by registration of multiple acoustic range views. In: Proceedings of the European Conference on Computer Vision. (2002) 805–819
15. Govindu, V.M.: Lie-algebraic averaging for globally consistent motion estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2004) 684–691
16. Torsello, A., Rodolà, E., Albarelli, A.: Multiview registration via graph diffusion of dual quaternions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2011) 2441 – 2448
17. Tron, R., Danilidis, K.: Statistical pose averaging with varying and non-isotropic covariances. In: Proceedings of the European Conference on Computer Vision. (2014)
18. Bernard, F., Thunberg, J., Gemmar, P., Hertel, F., Husch, A., Goncalves, J.: A solution for multi-alignment by transformation synchronisation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015)
19. Arrigoni, F., Rossi, B., Fusiello, A.: Spectral synchronization of multiple views in SE(3). SIAM Journal on Imaging Sciences **9**(4) (2016) 1963 – 1990
20. Arrigoni, F., Rossi, B., Fusiello, A.: Global registration of 3D point sets via LRS decomposition. In: Proceedings of the 14th European Conference on Computer Vision. (2016) 489–504
21. Rosen, D.M., DuHadway, C., Leonard, J.J.: A convex relaxation for approximate global optimization in simultaneous localization and mapping. In: Proceedings of the IEEE International Conference on Robotics and Automation. (2015) 5822 – 5829
22. Rosen, D., Carlone, L., Bandeira, A., Leonard, J.: A certifiably correct algorithm for synchronization over the special Euclidean group. In: International Workshop on the Algorithmic Foundations of Robotics (WAFR). (2016)
23. Pachauri, D., Kondor, R., Singh, V.: Solving the multi-way matching problem by permutation synchronization. In: Advances in Neural Information Processing Systems 26. (2013) 1860–1868
24. Shen, Y., Huang, Q., Srebro, N., Sanghavi, S.: Normalized spectral map synchronization. In Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R., eds.:

Advances in Neural Information Processing Systems 29. Curran Associates, Inc. (2016) 4925–4933
25. Yu, J.G., Xia, G.S., Samal, A., Tian, J.: Globally consistent correspondence of multiple feature sets using proximal Gauss–Seidel relaxation. Pattern Recognition **51** (2016) 255 – 267
26. Arrigoni, F., Maset, E., Fusiello, A.: Synchronization in the symmetric inverse semigroup. In: Proceedings of the International Conference on Image Analysis and Processing. (2017)
27. Arrigoni, F., Fusiello, A.: Bearing-based network localizability: A unifying view. IEEE Transactions on Pattern Analysis and Machine Intelligence (2018) 1–1
28. Maset, E., Arrigoni, F., Fusiello, A.: Practical and efficient multi-view matching. In: Proceedings of IEEE International Conference on Computer Vision. Volume 2. (2017) 5
29. Arrigoni, F., Rossi, B., Fragneto, P., Fusiello, A.: Robust rotation synchronization via low-rank and sparse matrix decomposition. CoRR **abs/1505.06079** (2015) Submitted to IJCV.
30. Arie-Nachimson, M., Kovalsky, S.Z., Kemelmacher-Shlizerman, I., Singer, A., Basri, R.: Global motion estimation from point matches. Proceedings of the Joint 3DIM/3DPVT Conference: 3D Imaging, Modeling, Processing, Visualization and Transmission (2012)
31. Singer, A., Shkolnisky, Y.: Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming. SIAM Journal on Imaging Sciences **4**(2) (2011) 543 – 572
32. Govindu, V.M.: Combining two-view constraints for motion estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2001)
33. Brand, M., Antone, M., Teller, S.: Spectral solution of large-scale extrinsic camera calibration as a graph embedding problem. In: Proceedings of the European Conference on Computer Vision. (2004)
34. Jiang, N., Cui, Z., Tan, P.: A global linear method for camera pose registration. In: Proceedings of the International Conference on Computer Vision. (2013)
35. Wilson, K., Snavely, N.: Robust global translations with 1DSfM. In: Proceedings of the European Conference on Computer Vision. (2014) 61–75
36. Goldstein, T., Hand, P., Lee, C., Voroninski, V., Soatto, S.: ShapeFit and ShapeKick for robust, scalable structure from motion. In: Proceedings of the European Conference on Computer Vision. (2016) 289 – 304
37. Arrigoni, F., Fusiello, A., Rossi, B.: On computing the translations norm in the epipolar graph. In: Proceedings of the International Conference on 3D Vision (3DV). (2015) 300–308
38. Arrigoni, F., Fusiello, A., Rossi, B.: Camera motion from group synchronization. In: Proceedings of the International Conference on 3D Vision (3DV). (2016) 546–555
39. Holland, P.W., Welsch, R.E.: Robust regression using iteratively reweighted least-squares. Communications in Statistics - Theory and Methods **6**(9) (1977) 813–827