

Supplementary Material

On the Usage of the Trifocal Tensor in Motion Segmentation

Federica Arrigoni¹, Luca Magri², and Tomas Pajdla³

¹ DISI, University of Trento, Italy – federica.arrigoni@unitn.it

² DEIB, Politecnico di Milano, Italy – luca.magri@polimi.it

³ CIIRC, CTU in Prague, Czech Republic – pajdla@cvut.cz

In this document we report technical derivations (Sec. 1) and visual results (Sec. 2) that could not be included in the main paper [1] due to space constraints.

1 Averaging permutation matrices

Let \mathcal{P} denote the set of $d \times d$ permutation matrices (also known as the *Symmetric Group*), namely $\mathcal{P} = \{P \in \{0, 1\}^{d \times d} \text{ s.t. } P\mathbf{1} = \mathbf{1}, \mathbf{1}P = \mathbf{1}\}$ where $\mathbf{1}$ is a vector of ones. A permutation matrix is such that exactly one entry in each row and column is equal to 1 and all other entries are zero. Given $Q_1, \dots, Q_r \in \mathcal{P}$, the task is to find a permutation $P \in \mathcal{P}$ – named the *mean* or *average* – that best represents the set $\{Q_i \text{ s.t. } i = 1, \dots, r\}$. A reasonable approach consists in addressing the following optimization problem

$$\min_{P \in \mathcal{P}} \sum_{i=1}^r \|P - Q_i\|_F^2 \quad (1)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. By computation we obtain

$$\sum_{i=1}^r \|P - Q_i\|_F^2 = \sum_{i=1}^r \left(\|P\|_F^2 + \|Q_i\|_F^2 - 2\text{trace}(P^T Q_i) \right) = 2rd - 2\text{trace}\left(P^T \sum_{i=1}^r Q_i\right) \quad (2)$$

where the squared Frobenius norm of any permutation matrix is equal to d . Thus

$$\min_{P \in \mathcal{P}} \sum_{i=1}^r \|P - Q_i\|_F^2 \iff \max_{P \in \mathcal{P}} \text{trace}\left(P^T \sum_{i=1}^r Q_i\right) \iff \min_{P \in \mathcal{P}} \|P - \sum_{i=1}^r Q_i\|_F^2 \quad (3)$$

meaning that solving Problem (1) is equivalent to find the closest permutation matrix to the sum of the input permutations. Such a task is a linear assignment problem [5]. Observe that this procedure resembles the “chordal L_2 -mean” of a set of rotation matrices [4].

2 Qualitative results

Our dataset comprises six indoor scenes with seven images, which are reported in Fig. 1, 2, 3, 4, 5 and 6. Each scene counts a number of moving objects captured by a moving camera. One additional object is not moving, meaning that it is part of the background. Two-frame correspondences were obtained with SIFT [6] and ground-truth segmentation was established by manually creating contours around each object. Segmentation results obtained by our methods (namely TRISEG and TRIPAIRSEG) and by competing techniques (namely MODE [3] and SYNCH [2]) are reported in Fig. 1, 2, 3, 4, 5 and 6. Both TRISEG and TRIPAIRSEG provide accurate results, where the former returns a very clean segmentation with fewer points. Observe that most of the unclassified points belong to the background, which is textureless. SYNCH successfully solves motion segmentation only in the `stuffed_animals5` sequence (see Fig. 5) and it fails in the remaining cases. MODE provides very good results in sequences with three motions (see Fig. 5 and 6) whereas it has difficulties in handling all those scenes with four motions, to different extends. For example, observe that MODE oversegments one object in the fourth image in Fig. 1c and the same happens in Fig. 2c. Moreover, it gives the wrong label to an entire object in the second image in Fig. 3c and it permutes the labels of two objects in the fifth image in Fig. 4c. This analysis confirms the quantitative evaluation carried out in the main paper.

References

1. Arrigoni, F., Magri, L., Pajdla, T.: On the usage of the trifocal tensor in motion segmentation. In: Proceedings of the European Conference on Computer Vision (2020)
2. Arrigoni, F., Pajdla, T.: Motion segmentation via synchronization. In: IEEE International Conference on Computer Vision Workshops (ICCVW) (2019)
3. Arrigoni, F., Pajdla, T.: Robust motion segmentation from pairwise matches. In: Proceedings of the International Conference on Computer Vision (2019)
4. Hartley, R.I., Trunpf, J., Dai, Y., Li, H.: Rotation averaging. International Journal of Computer Vision (2013)
5. Kuhn, H.W.: The Hungarian method for the assignment problem. Naval Research Logistics Quarterly 2 **2**, 83 – 97 (1955)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2), 91–110 (2004). <https://doi.org/http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>

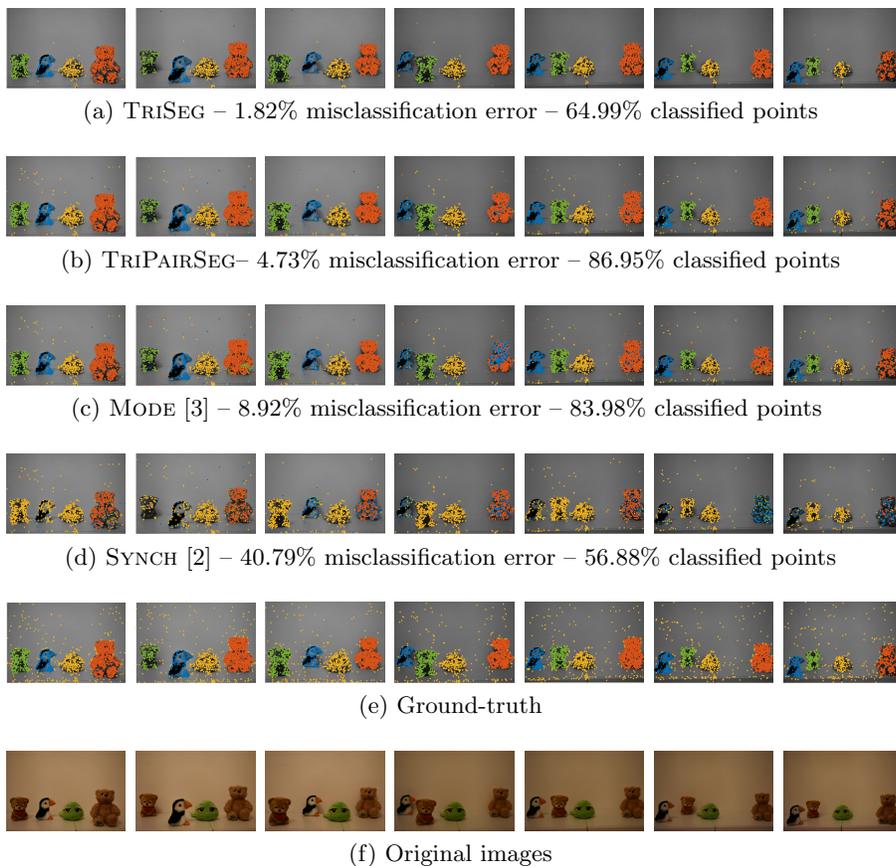


Fig. 1: This figure reports qualitative results obtained by several methods on the `stuffed_animals1` sequence. Different colours correspond to different motions and unclassified points are not drawn. Ground-truth segmentation and original images are also reported.

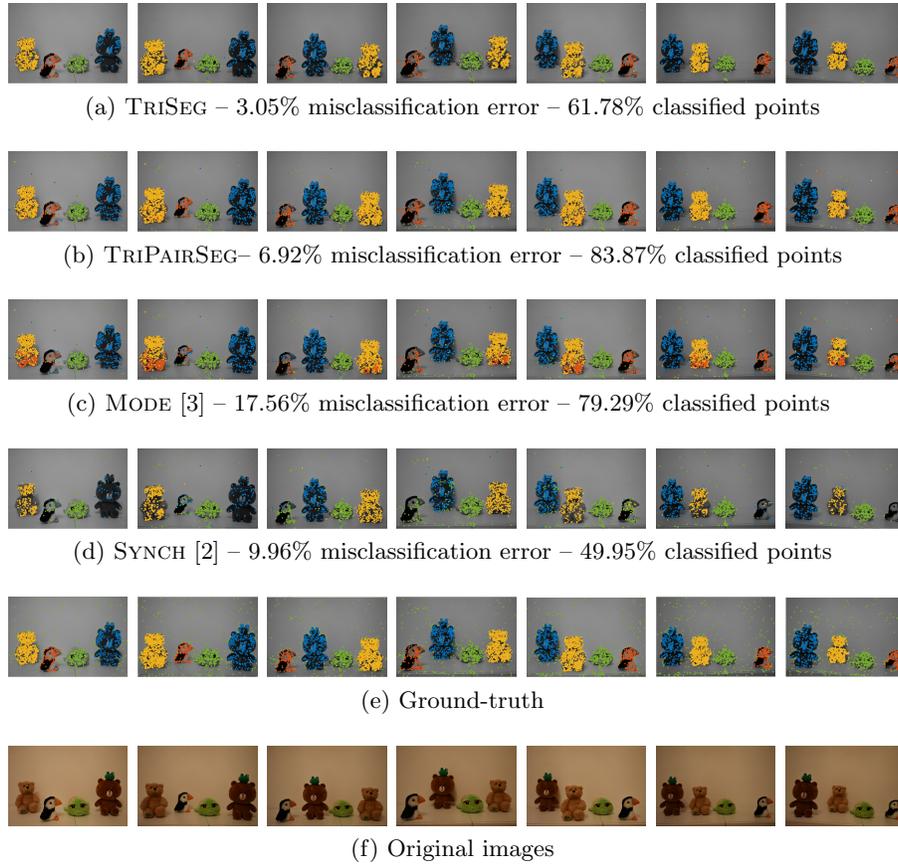


Fig. 2: This figure reports qualitative results obtained by several methods on the `stuffed_animals2` sequence. Different colours correspond to different motions and unclassified points are not drawn. Ground-truth segmentation and original images are also reported.



Fig. 3: This figure reports qualitative results obtained by several methods on the `stuffed_animals3` sequence. Different colours correspond to different motions and unclassified points are not drawn. Ground-truth segmentation and original images are also reported.

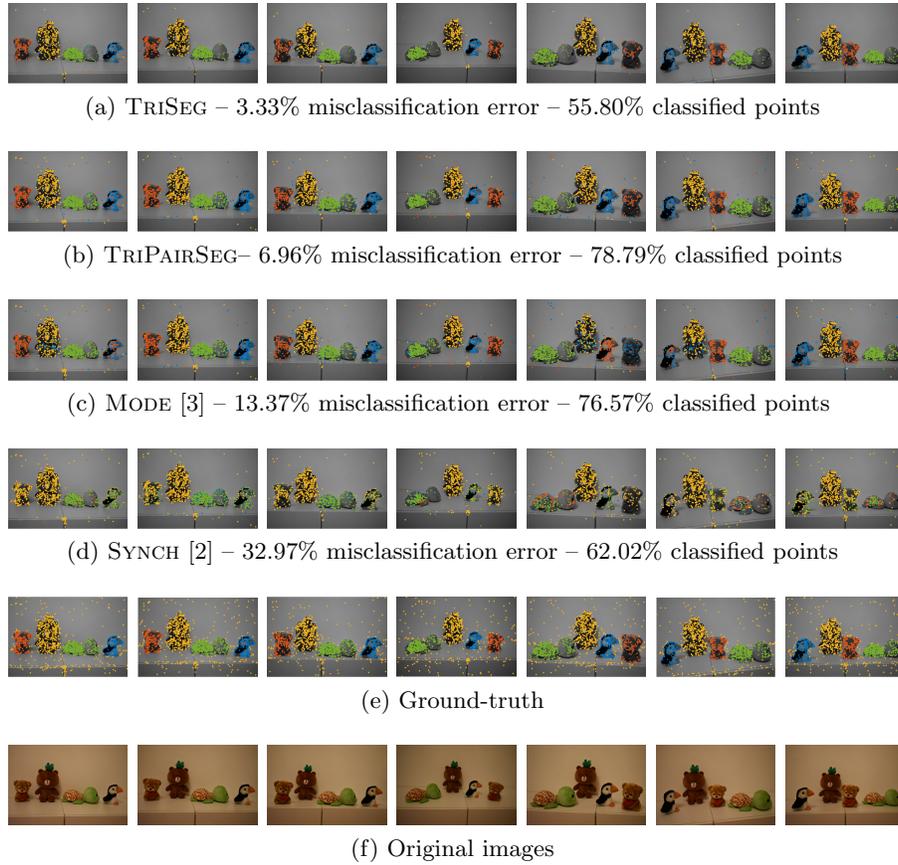


Fig. 4: This figure reports qualitative results obtained by several methods on the `stuffed_animals4` sequence. Different colours correspond to different motions and unclassified points are not drawn. Ground-truth segmentation and original images are also reported.

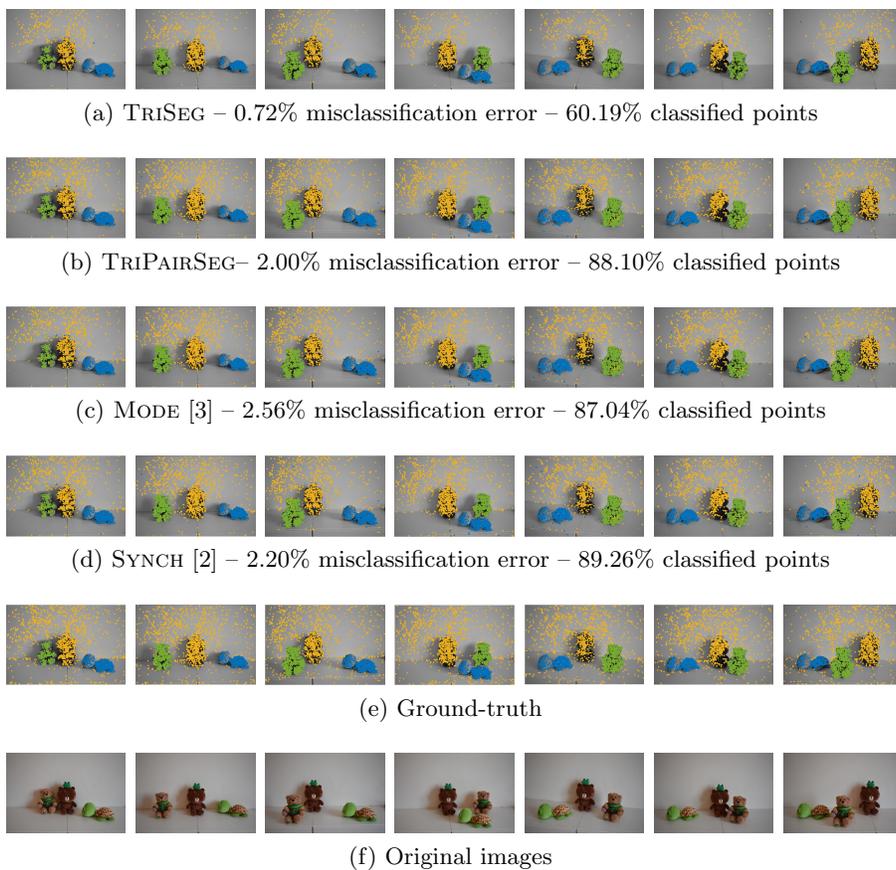


Fig. 5: This figure reports qualitative results obtained by several methods on the `stuffed_animals5` sequence. Different colours correspond to different motions and unclassified points are not drawn. Ground-truth segmentation and original images are also reported.

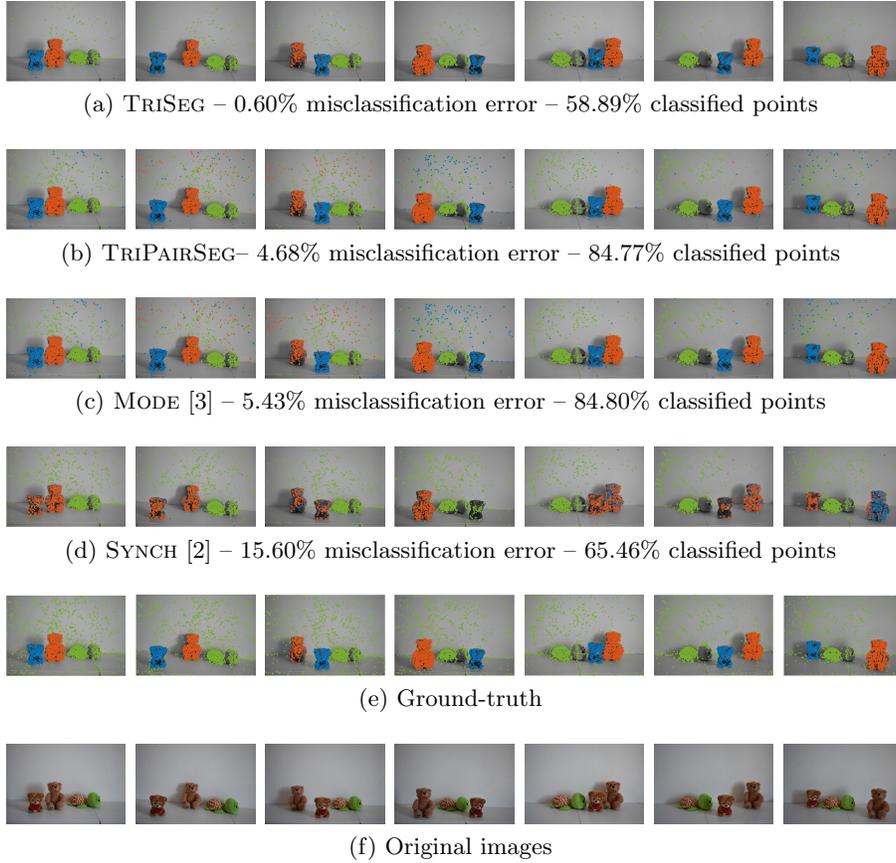


Fig. 6: This figure reports qualitative results obtained by several methods on the `stuffed_animals6` sequence. Different colours correspond to different motions and unclassified points are not drawn. Ground-truth segmentation and original images are also reported.