Supplementary Material
Robust Motion Segmentation From Pairwise Matches

Federica Arrigoni and Tomas Pajdla
CIIRC – Czech Technical University in Prague
Federica.Arrigoni@cvut.cz, pajdla@cvut.cz

1. Synthetic Experiments

We provide additional analysis illustrating what happens to individual points when running our method. We refer here to the experimental results reported in the main paper where the *cars1* sequence [7] was used and synthetic errors were added to pairwise correspondences. Figure 2 reports coloured bars representing the amount of errors for each point in a sample image. As the percentage of mismatches increases, motion segmentation gets harder to solve, since the green area reduces whereas the blue and red ones enlarge. Note that RPA [5] produces errors even in the absence of wrong correspondences, as can be appreciated in Fig. 2a. Our method classifies all the data except for a few cases where the blue bars are equal to 1, meaning that the point is labelled as outlier by RPA in all the pairs. Among the classified points, MODE provides a correct segmentation as long as the green bars are sufficiently high.

2. Real Experiments

2.1. Indoor scenes

Our benchmark consists of five sequences of indoor scenes with two or three motions, which are shown in Fig. 4, 5, 6, 7 and 8. Observe that in the case of the *Penguin* sequence there is no motion between some images, namely frames 1 and 2, frames 3 and 4, frames 5 and 6. SIFT keypoints [4] were extracted in all the images and correspondences between image pairs were established using nearest neighbor and ratio test as in [4]. For each image pair (i, j), we kept only those correspondences that were found both when matching image i with j and when matching image j with i, and isolated features (i.e. points that are not matched in any image) were removed. No further filtering was applied. The properties of each sequence are summarized in Tab. 1. Ground-truth segmentation was established by manually labelling points in each image. The number of points that undergo the same motion is reported in Fig. 1, which gives an idea about the distribution of points in the scene for each sequence.

Table 1: The number of motions d, the number of images n, and the total number of image points p are reported for each sequence in our dataset.

Dataset	d	n	p
Penguin	2	6	5865
Flowers	2	6	7743
Pencils	2	6	2982
Bag	2	7	6114
Bears	3	10	15888

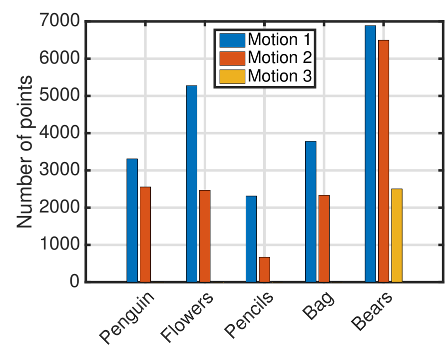
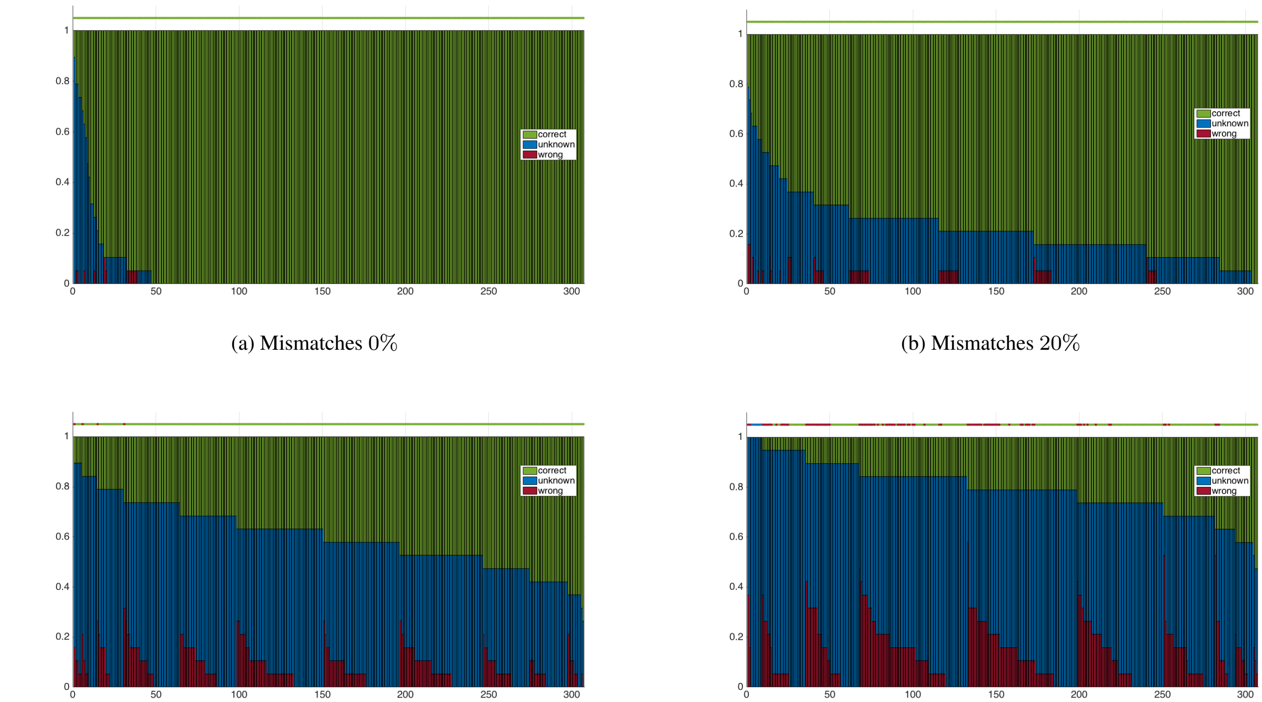


Figure 1: The number of points per motion is reported for each sequence in our dataset.

Figures 4, 5, 6, 7 and 8 visually represent the segmentation of image points obtained by several methods, which complement the quantitative evaluation provided in the main paper. Ground-truth segmentation is also shown. Concerning the different variants of Subset [9] and RSIM [2], which differ for the algorithm used for computing tracks, we report results for StableSfM [6] only. Indeed, there are not significative differences between StableSfM [6] and QuichMatch [8] in terms of misclassification error, but the former is better in terms of amount of classified data. Our method returns high quality (although not perfect) segmentation in all the sequences, outperforming the baseline in terms of percentage of classified points, whereas Subset and RSIM exhibit poor performances in our dataset.



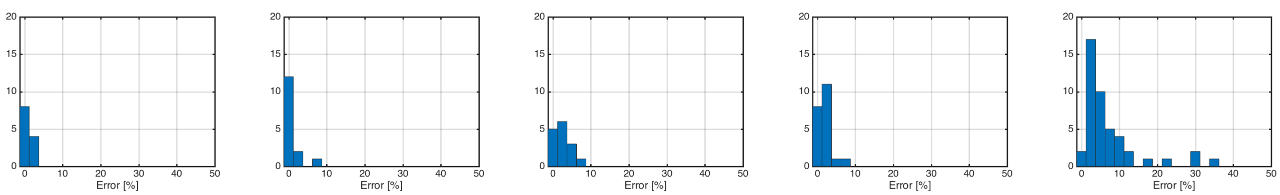
(a) Mismatches 0%

(b) Mismatches 20%

(c) Mismatches 60%

(d) Mismatches 80%

Figure 2: The horizontal axis indexes points in a sample image from *cars1* [7] and a three-color bar is shown for each point. Bars are divided into three parts which sum to one. The green, red, and blue parts represent fractions of image pairs where the point is correctly classified, misclassified, and labeled as outlier, respectively, by RPA [5]. For better visualization, points are sorted increasingly by the height of green bars. A dot is plotted over each bar to show whether the point is classified by our method correctly (green), misclassified (red) or labelled as unknown (blue).



(a) Penguin

(b) Flowers

(c) Pencils

(d) Bag

(e) Bears

Figure 3: Histograms of misclassification error achieved by RPA [5] on all the sequences from our dataset. Each point in the horizontal axis corresponds to a possible misclassification error in an individual image pair, and each point in the vertical axis corresponds to the number of pairs where a given error is reached.

In order to give further insights on the behavior of our technique, we report in Fig. 3 the histograms of misclassification error achieved by RPA [5] over image pairs, similarly to the synthetic experiments conducted in the main paper. The histograms show the effective amount of corruption in the data after performing pairwise segmentation with RPA, which is the first step of our pipeline. Note that the misclassification error exceeds 30% in some image pairs from the *Bears* sequence. It is remarkable that our method is able to achieve a low error in this dataset (about 4.8%), as reported in the main paper. In other words, it can effectively reduce errors in the pairwise segmentations thanks to the fact that it exploits redundant measures.

2.2. Outdoor scenes

Figures 9, 10, 11 and 12 visually report the segmentation obtained by MODE, the baseline and Subset [9] combined with StableSfM [6] on all the images of the considered outdoor scenes, namely *helicopter* [1], *boat* [3], *cars7* [7] and *cars8* [7], that were not included in the main paper due to space constraints. Our method provides high quality segmentation on all the sequences, outperforming the baseline in terms of percentage of classified data. While there are no significant differences between MODE and Subset in the *boat* sequence, the improvement of our method over the latter is evident in the *helicopter*, *cars7* and *cars8* sequences.

References

[1] Ralf Dragon, Jörn Ostermann, and Luc Van Gool. Robust realtime motion-split-and-merge for motion segmentation. In *German Conference on Pattern Recognition*, pages 425–434. Springer Berlin Heidelberg, 2013.

[2] Pan Ji, Mathieu Salzmann, and Hongdong Li. Shape interaction matrix revisited and robustified: Efficient subspace clustering with corrupted and incomplete data. In *Proceedings of the International Conference on Computer Vision*, pages 4687–4695, 2015.

[3] Zhuwen Li, Jiaming Guo, Loong-Fah Cheong, and Steven Zhiying Zhou. Perspective motion segmentation via collaborative clustering. In *Proceedings of the International Conference on Computer Vision*, pages 1369–1376, 2013.

[4] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

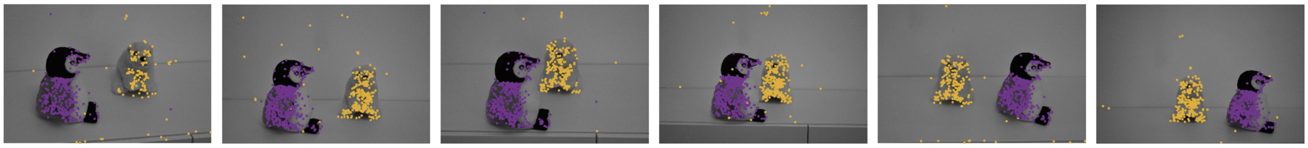
[5] Luca Magri and Andrea Fusiello. Robust multiple model fitting with preference analysis and low-rank approximation. In *Proceedings of the British Machine Vision Conference*, pages 20.1–20.12. BMVA Press, September 2015.

[6] Carl Olsson and Olof Enqvist. Stable structure from motion for unordered image collections. In *Proceedings of the 17th Scandinavian conference on Image analysis (SCIA'11)*, pages 524–535. Springer-Verlag, 2011.

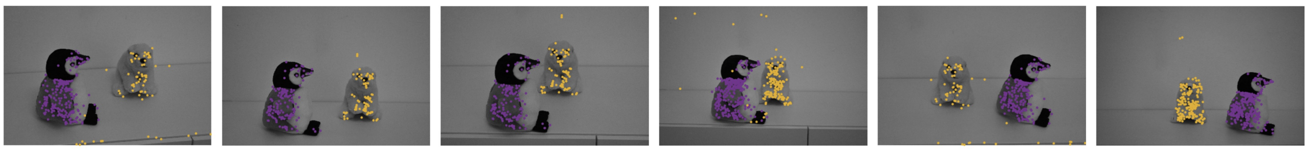
[7] Roberto Tron and René Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[8] Roberto Tron, Xiaowei Zhou, Carlos Esteves, and Kostas Daniilidis. Fast multi-image matching via density-based clustering. In *Proceedings of the International Conference on Computer Vision*, pages 4077–4086, 2017.

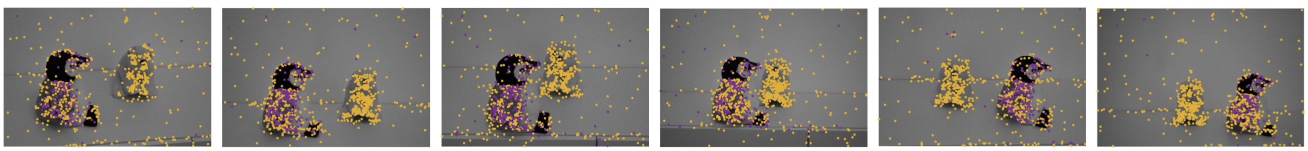
[9] Xun Xu, Loong-Fah Cheong, and Zhuwen Li. Motion segmentation by exploiting complementary geometric models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.



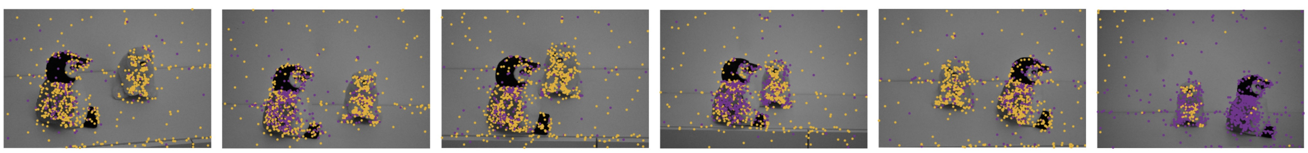
(a) MODE – 0.76% misclassification error – 69.17% classified points



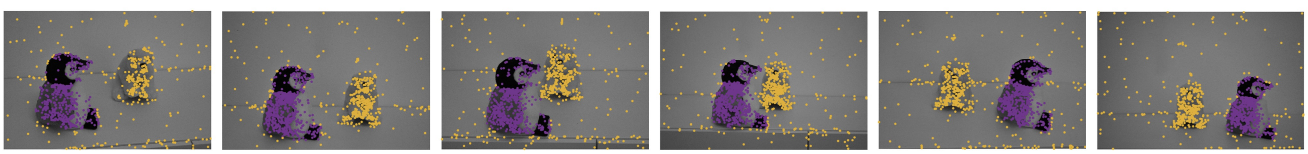
(b) Baseline – 0.95% misclassification error – 33.95% classified points



(c) Subset – 32.27% misclassification error – 99.59% classified points



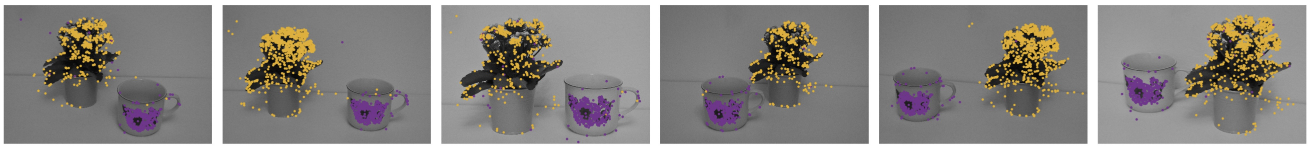
(d) RSIM – 41.50% misclassification error – 99.59% classified points



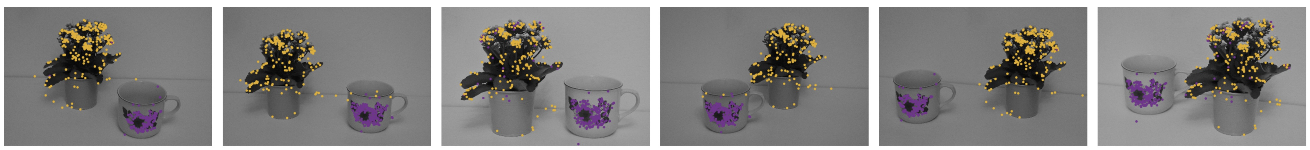
(e) Ground-truth

(f) Original images

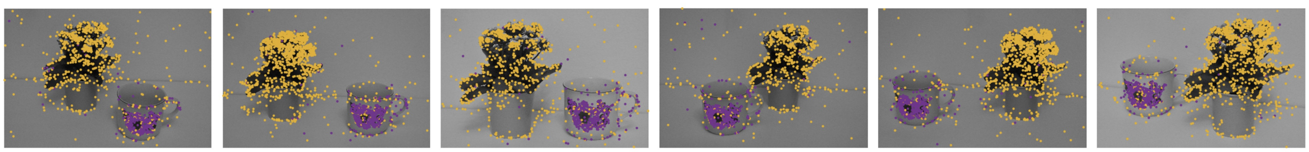
Figure 4: Segmentation results are reported for several methods on the *Penguin* sequence. Images are drawn in grey-scale and points are drawn in different colors based on the membership to different motions. For better visualization, unclassified points are not drawn. Ground-truth segmentation is also reported, in addition to original (coloured) images.



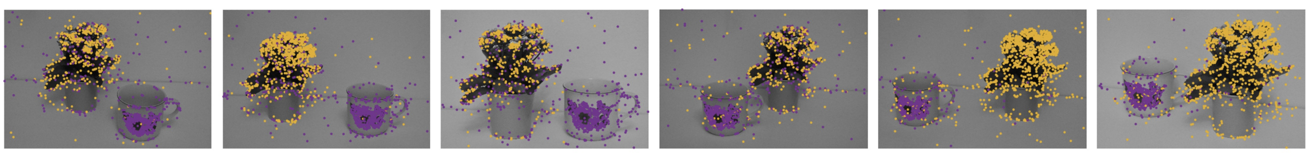
(a) MODE – 1.23% misclassification error – 73.65% classified points



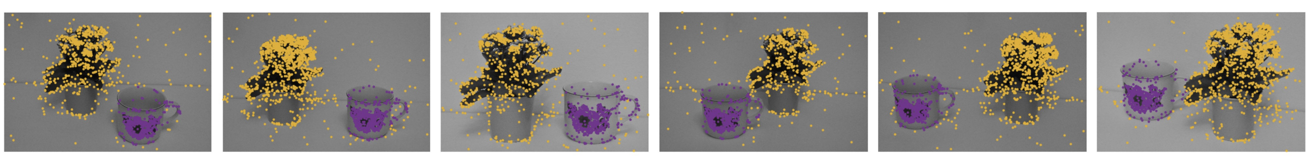
(b) Baseline – 2.84% misclassification error – 32.70% classified points



(c) Subset – 8.55% misclassification error – 99.50% classified points



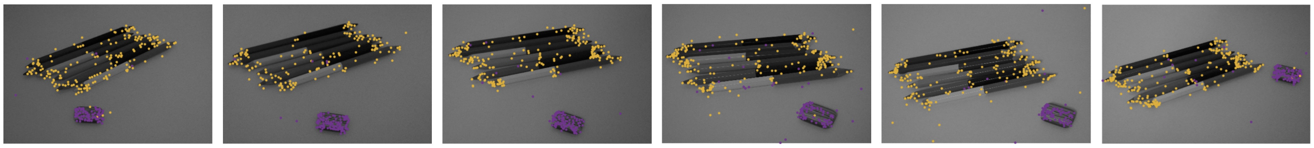
(d) RSIM – 16.65% misclassification error – 99.50% classified points



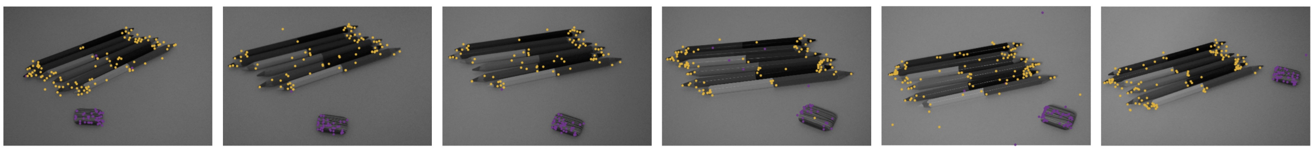
(e) Ground-truth

(f) Original images

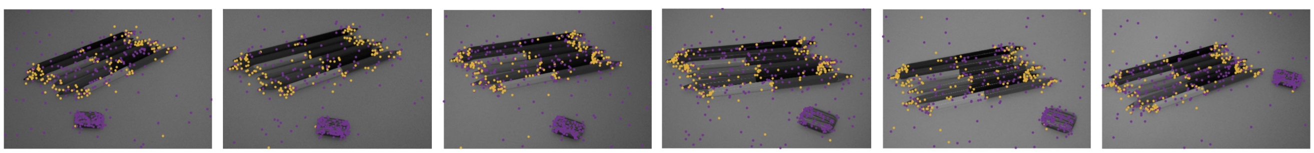
Figure 5: Segmentation results are reported for several methods on the *Flowers* sequence. Images are drawn in grey-scale and points are drawn in different colors based on the membership to different motions. For better visualization, unclassified points are not drawn. Ground-truth segmentation is also reported, in addition to original (coloured) images.



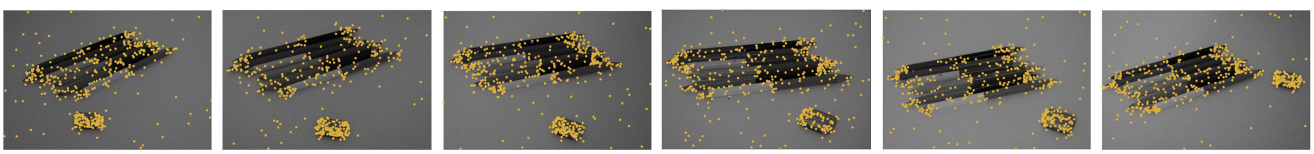
(a) MODE – 3.80% misclassification error – 65.33% classified points



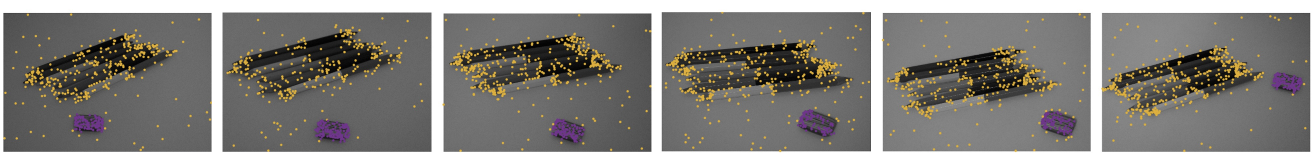
(b) Baseline – 2.30% misclassification error – 30.65% classified points



(c) Subset – 41.46% misclassification error – 99.56% classified points



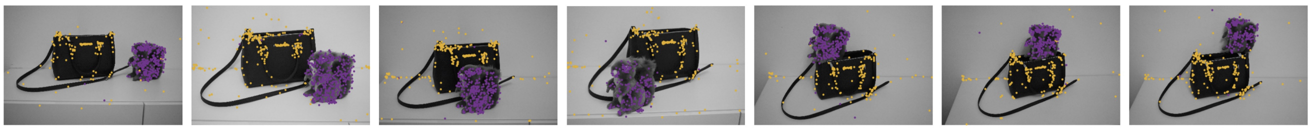
(d) RSIM – 23.07% misclassification error – 99.56% classified points



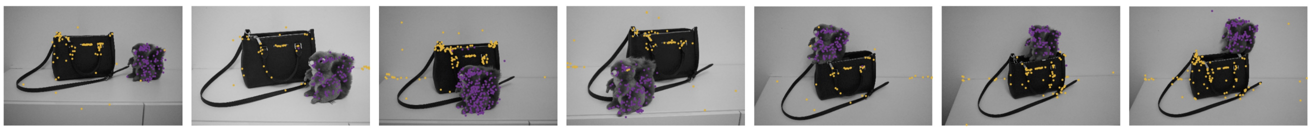
(e) Ground-truth

(f) Original images

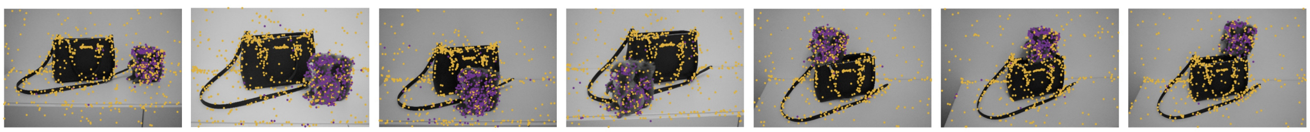
Figure 6: Segmentation results are reported for several methods on the *Pencils* sequence. Images are drawn in grey-scale and points are drawn in different colors based on the membership to different motions. For better visualization, unclassified points are not drawn. Ground-truth segmentation is also reported, in addition to original (coloured) images.



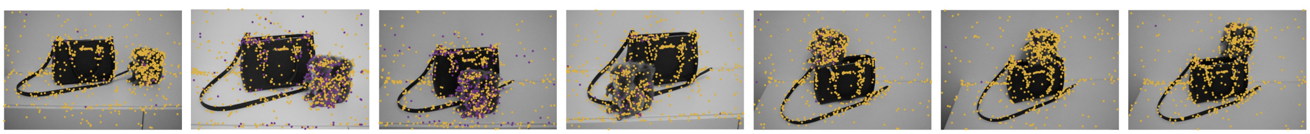
(a) MODE – 1.52% misclassification error – 57.95% classified points



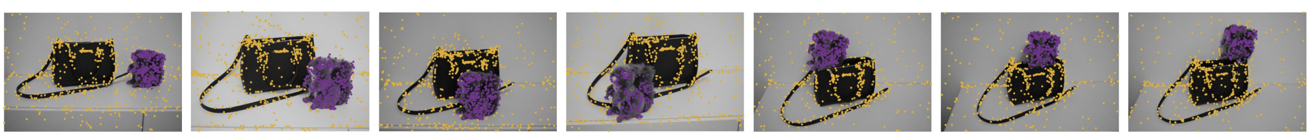
(b) Baseline – 1.54% misclassification error – 26.56% classified points



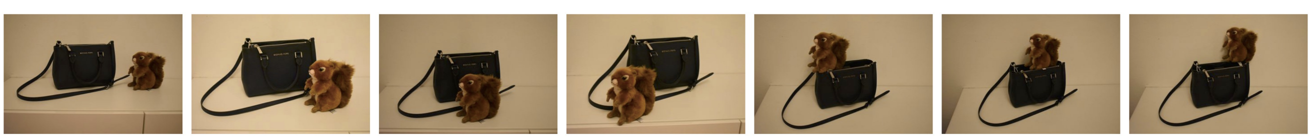
(c) Subset – 14.22% misclassification error – 99.69% classified points



(d) RSIM – 34.55% misclassification error – 99.69% classified points

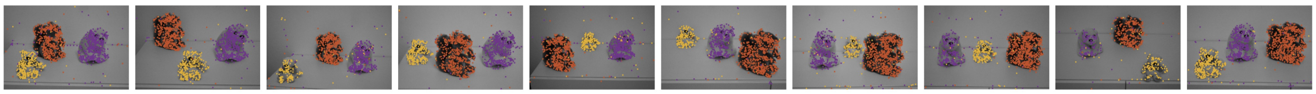


(e) Ground-truth

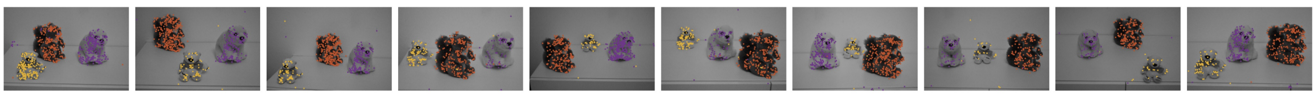


(f) Original images

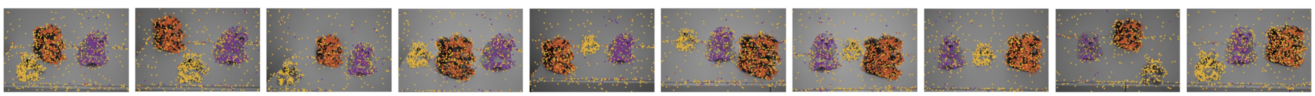
Figure 7: Segmentation results are reported for several methods on the *Bag* sequence. Images are drawn in grey-scale and points are drawn in different colors based on the membership to different motions. For better visualization, unclassified points are not drawn. Ground-truth segmentation is also reported, in addition to original (coloured) images.



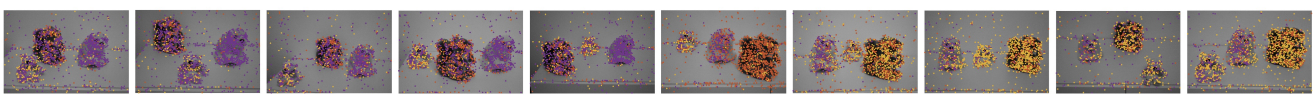
(a) MODE – 4.82% misclassification error – 73.65% classified points



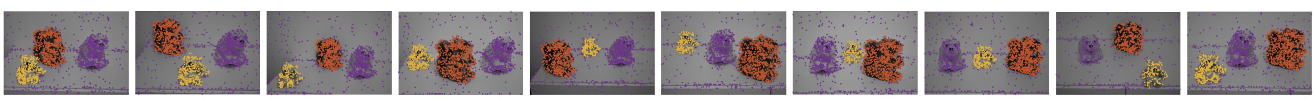
(b) Baseline – 2.72% misclassification error – 29.80% classified points



(c) Subset – 38.13% misclassification error – 99.58% classified points



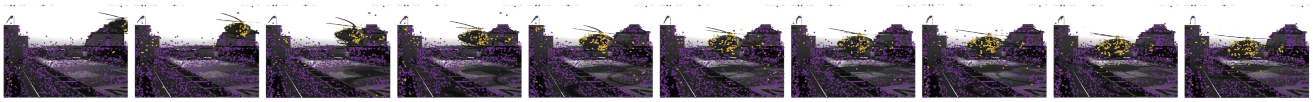
(d) RSIM – 49.48% misclassification error – 99.58% classified points



(e) Ground-truth

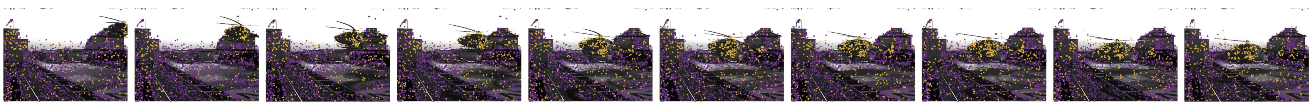
(f) Original images

Figure 8: Segmentation results are reported for several methods on the *Bears* sequence. Images are drawn in grey-scale and points are drawn in different colors based on the membership to different motions. For better visualization, unclassified points are not drawn. Ground-truth segmentation is also reported, in addition to original (coloured) images.

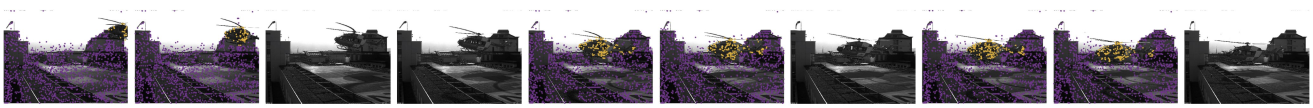


(a) MODE – 2.01% misclassification error – 80.82% classified points

(b) Baseline – 0.78% misclassification error – 45.93% classified points



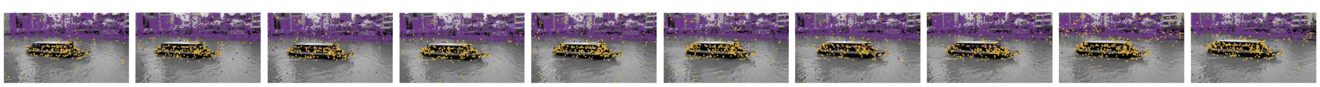
(c) Subset – 16.81% misclassification error – 99.52% classified points



(d) Ground-truth

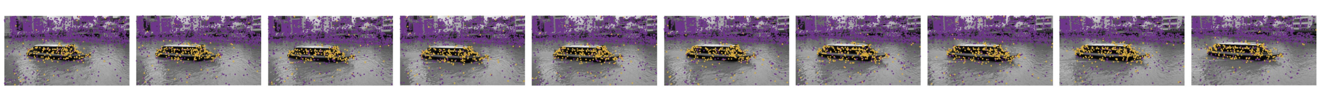
(e) Original images

Figure 9: Segmentation results are reported for several methods on the *helicopter* sequence [1]. Images are drawn in grey-scale and points are drawn in different colors based on the membership to different motions. For better visualization, unclassified points are not drawn. Ground-truth segmentation is reported only for those images for which ground-truth pixel-wise annotation is provided. Original (coloured) images are also reported.

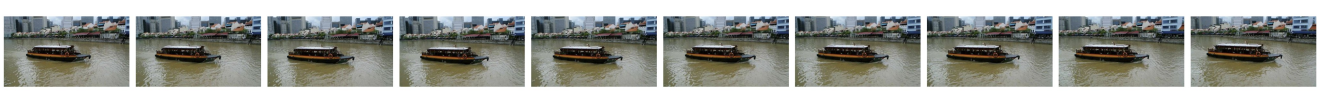


(a) MODE – 87.34% classified points

(b) Baseline – 56.31% classified points



(c) Subset – 99.62% classified points

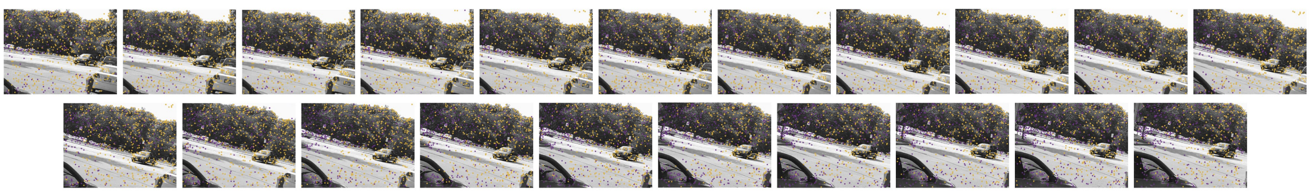


(d) Original images

Figure 10: Segmentation results are reported for several methods on the *boat* sequence [3]. Images are drawn in grey-scale and points are drawn in different colors based on the membership to different motions. For better visualization, unclassified points are not drawn. Original (coloured) images are also reported.

(a) MODE – 92.27% classified points

(b) Baseline – 57.38% classified points



(c) Subset – 99.66% classified points

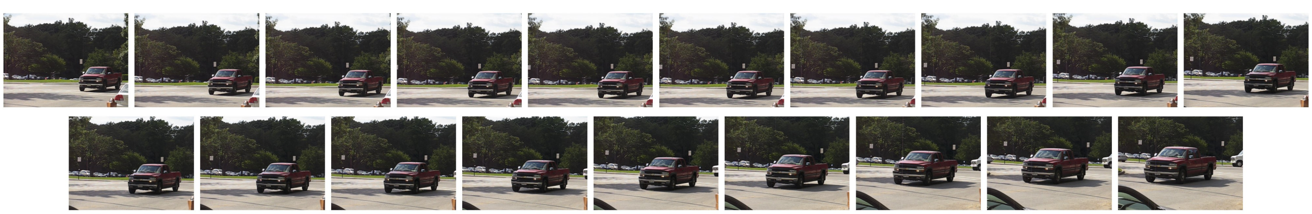
(d) Original images

Figure 11: Segmentation results are reported for several methods on the *cars7* sequence [7]. Images are drawn in grey-scale and points are drawn in different colors based on the membership to different motions. For better visualization, unclassified points are not drawn. Original (coloured) images are also reported.

(a) MODE – 93.12% classified points

(b) Baseline – 50.53% classified points

(c) Subset – 99.61% classified points



(d) Original images

Figure 12: Segmentation results are reported for several methods on the *cars8* sequence [7]. Images are drawn in grey-scale and points are drawn in different colors based on the membership to different motions. For better visualization, unclassified points are not drawn. Original (coloured) images are also reported.