

Segmentation of Dynamic Scenes with Distributions of Spatiotemporally Oriented Energies

Supplementary material

This document provides additional details and results of our method for video segmentation.

1 Comparison with optical flow

In the main paper, Fig. 1 right (reproduced below), we show a comparison between the optical flow and our peak motion energies. These figures use the classical representation of optical flow, *i.e.* the hue and saturation of a pixel are set respectively from the flow orientation and magnitude. We use a similar representation for the peak of our motion energies, so as to highlight the similarity with the optical flow. The hue and saturation of each pixel are set respectively from the image (spatial) orientation and (spatiotemporal) velocity of the \hat{n}_i giving the maximum $ME_{\hat{n}_i}$ at that pixel.

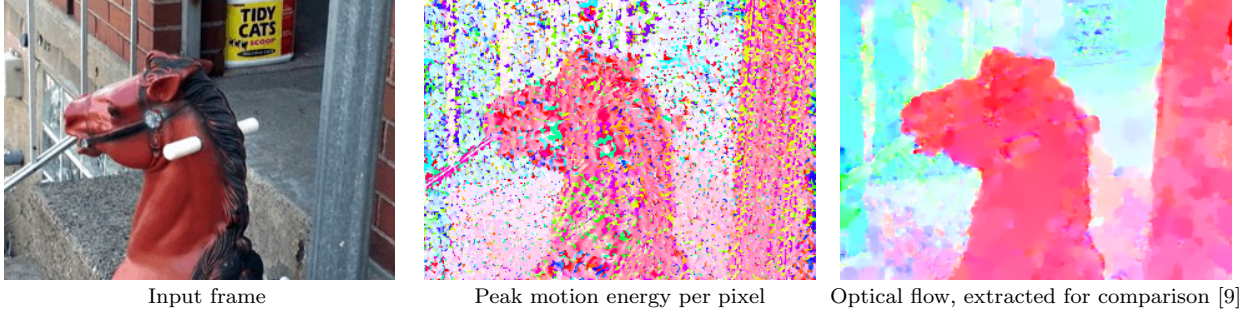


Figure 1: Comparison between the peak motion energy $ME_{\hat{n}_i}$ at each pixel and optical flow.

2 Dynamic texture segmentation

The reported quantitative results on the SynthDB dataset use the level of the segmentation that gives the same number of segments as the ground truth (2, 3, or 4). In other words the number of segments is known. To our knowledge, the results reported for the LDT [2] and DTM [1] methods use static segments, whereas the $(LBP/WLD)_{TOP}$ [3] method produces moving segments. We produced both to allow for a fair comparison with all methods. Enforcing static segments obviously makes the segmentation easier.

A video of the segmented “ocean-fire” sequence is available as supplementary material and on the authors’ website.

3 Motion segmentation

The sequences of the MIT dataset contain varying numbers of segments. Our method produces different levels of segmentation, and we therefore report below the performance at different levels, *i.e.* for different numbers of segments. The agreement with the ground truth is measured with the Rand index (ratio of pairwise agreements between pairs of pixels), which does not require an equal number of segments in the ground truth and in the result. Since, in practice, the number of segments in each scene is not necessarily known, and since several of the levels of segmentation produced are equally plausible interpretations of the scenes (at different levels of granularity), the quantitative results reported in the main paper correspond to the level giving the best performance for each sequence.

The segments are colored randomly for visualization purposes. The boundary between two segments is colored with the same hue as the segment it is assigned to, according to our method for depth ordering. The color of boundaries for the “color only” evaluation is irrelevant.

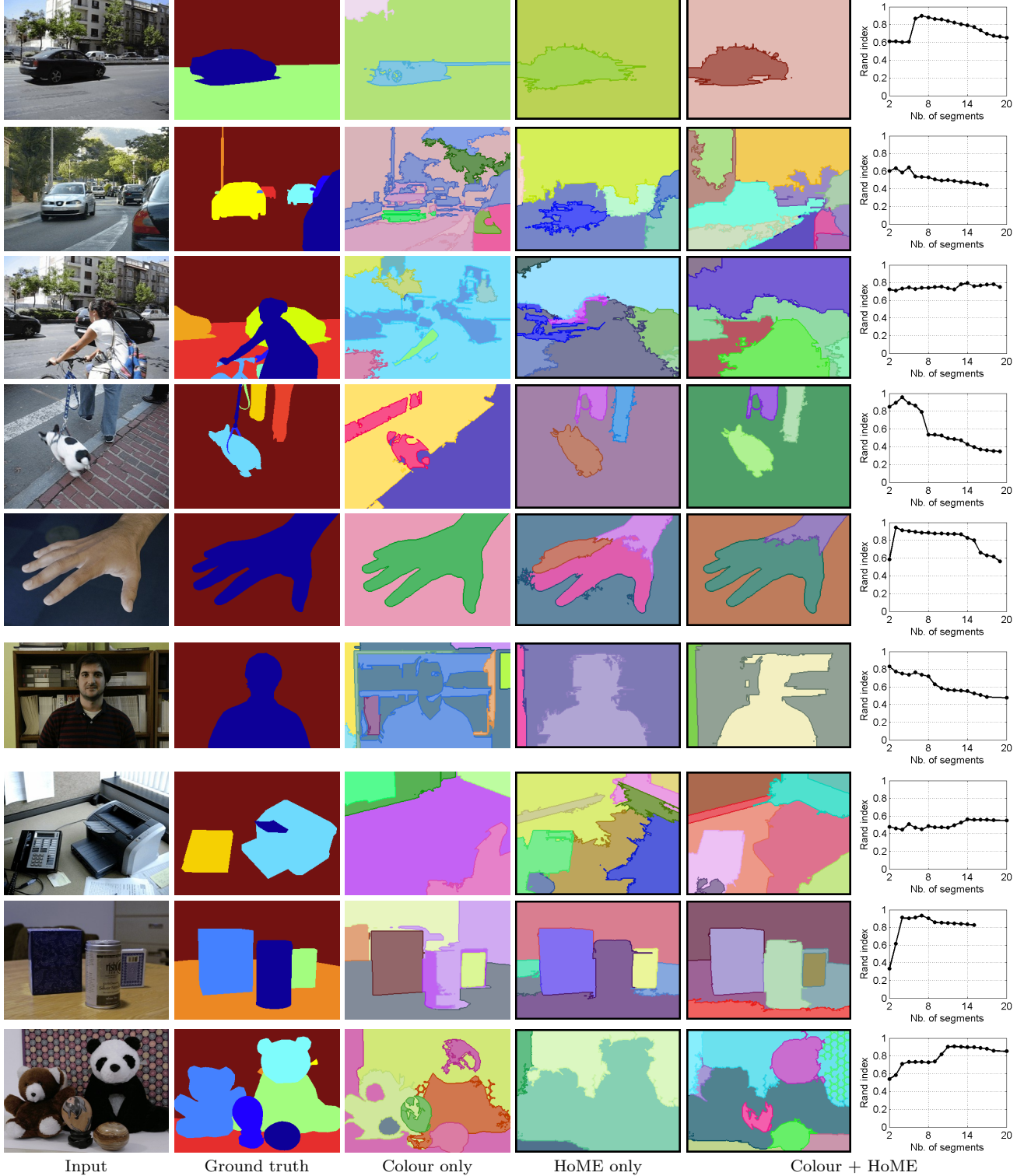


Figure 2: Motion segmentation on the MIT dataset.

4 Occlusion boundaries

We use the CMU dataset to evaluate our capability of detecting occlusion boundaries. This dataset of 30 short sequences features mostly static objects, with slight camera translations resulting in relative motions due to parallax at different depths. A number of methods have been evaluated on this dataset, though in different conditions and with different protocols [7, 4, 6, 8, 5]. Most methods separate the tasks of detecting edges, and classifying them as boundaries or interior edges. The hand-drawn ground truth, provided with the dataset, is “fitted” to the candidate edges, and the results serve as the actual ground truth for the evaluation of these methods. Such an evaluation is not possible in our case, since detection and classification are not separate. We rather compare our boundaries directly against the hand-drawn ground truth. The boundaries of Stein and Hebert [7] were the only ones available publicly, and we thus evaluated them with our protocol; this should ensure a fair comparison with our method. Boundaries were counted as correct as per the usual criterion of the Berkeley segmentation benchmark, *i.e.* if they fall within 1% of the image diagonal to the ground truth (hand drawn boundaries).

We provide below our results on all the 30 sequences of the dataset. The precision/recall curves correspond to the detection of boundaries.

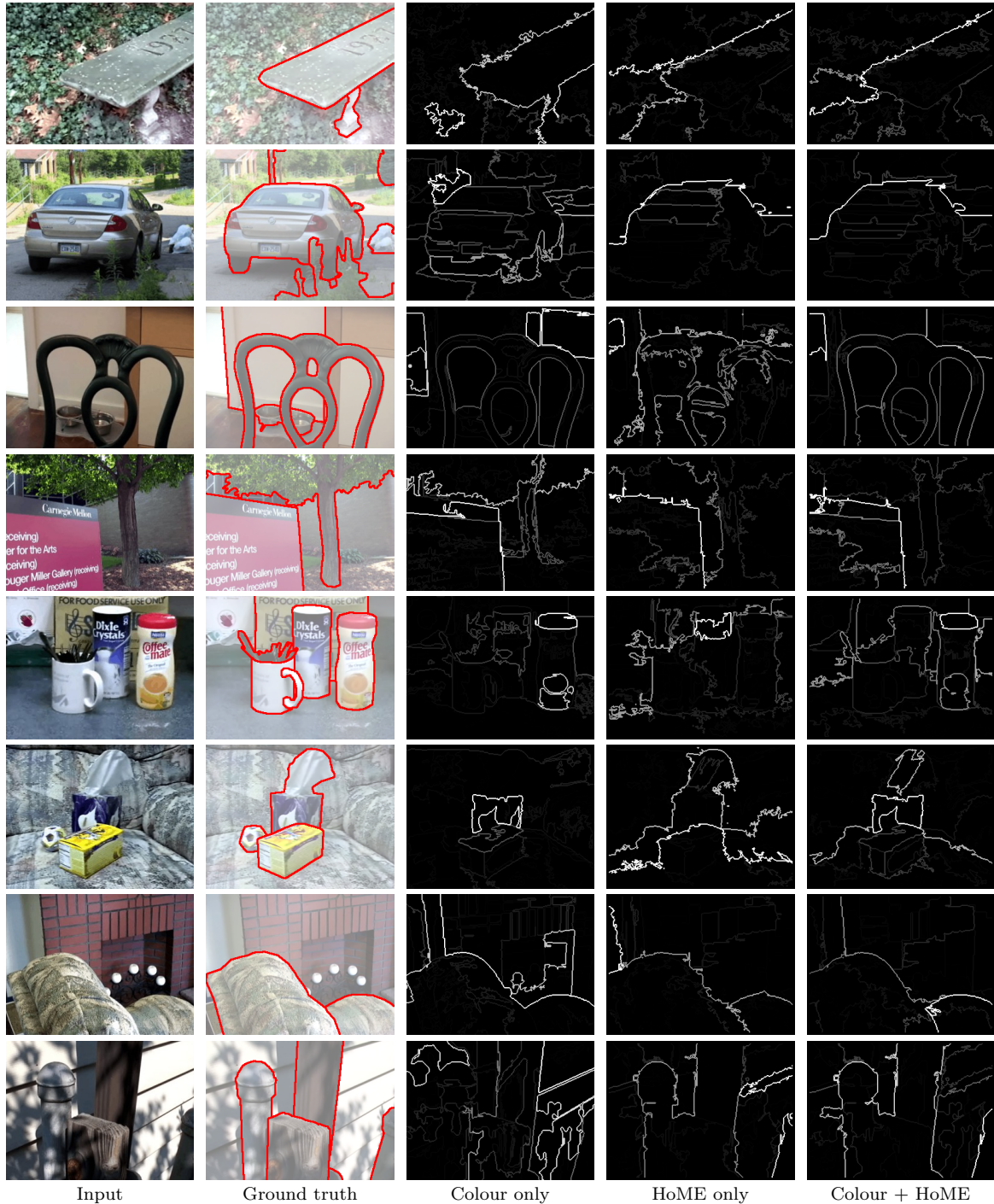


Figure 3: Results on the CMU dataset.

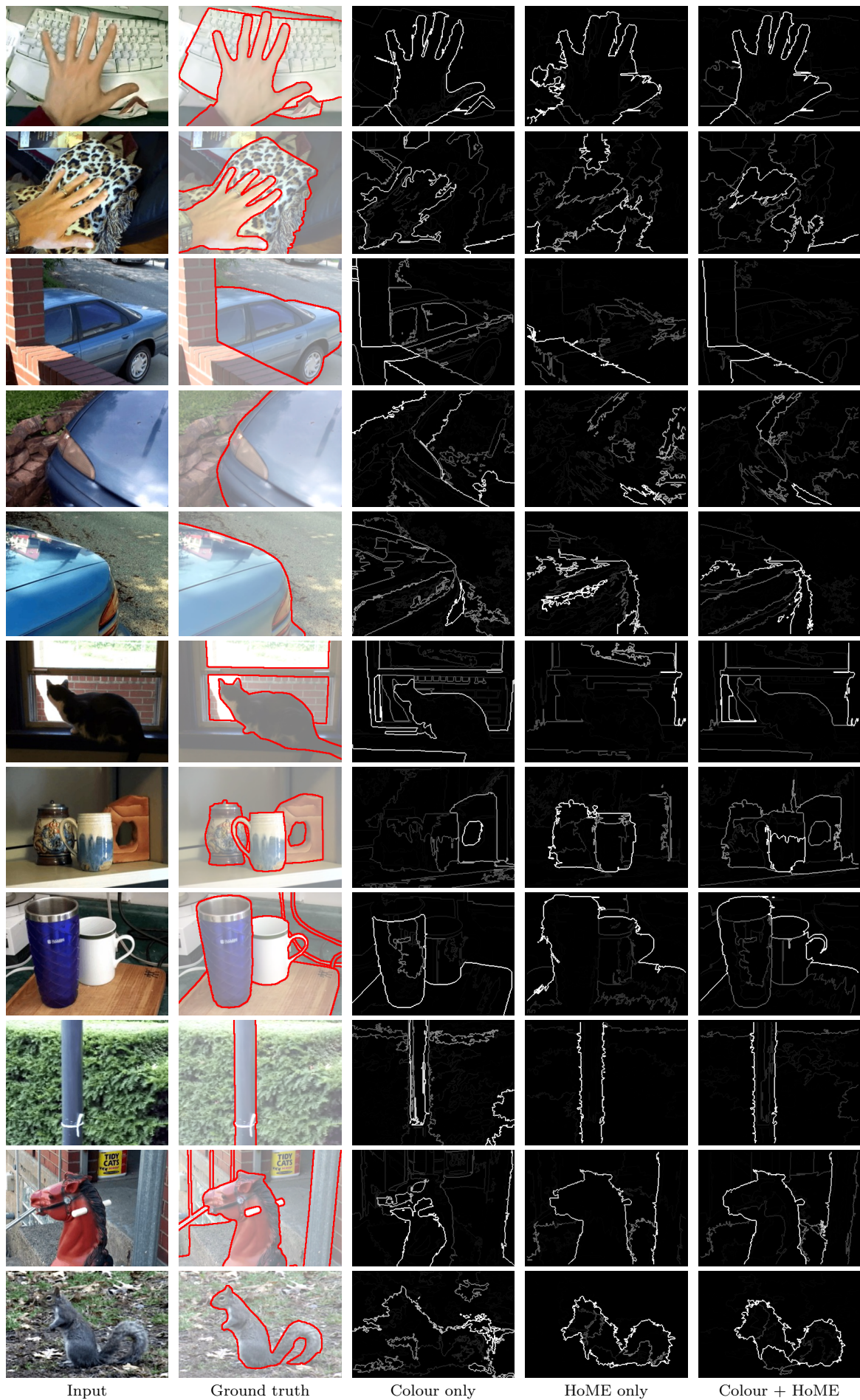


Figure 4: Results on the CMU dataset (continued).

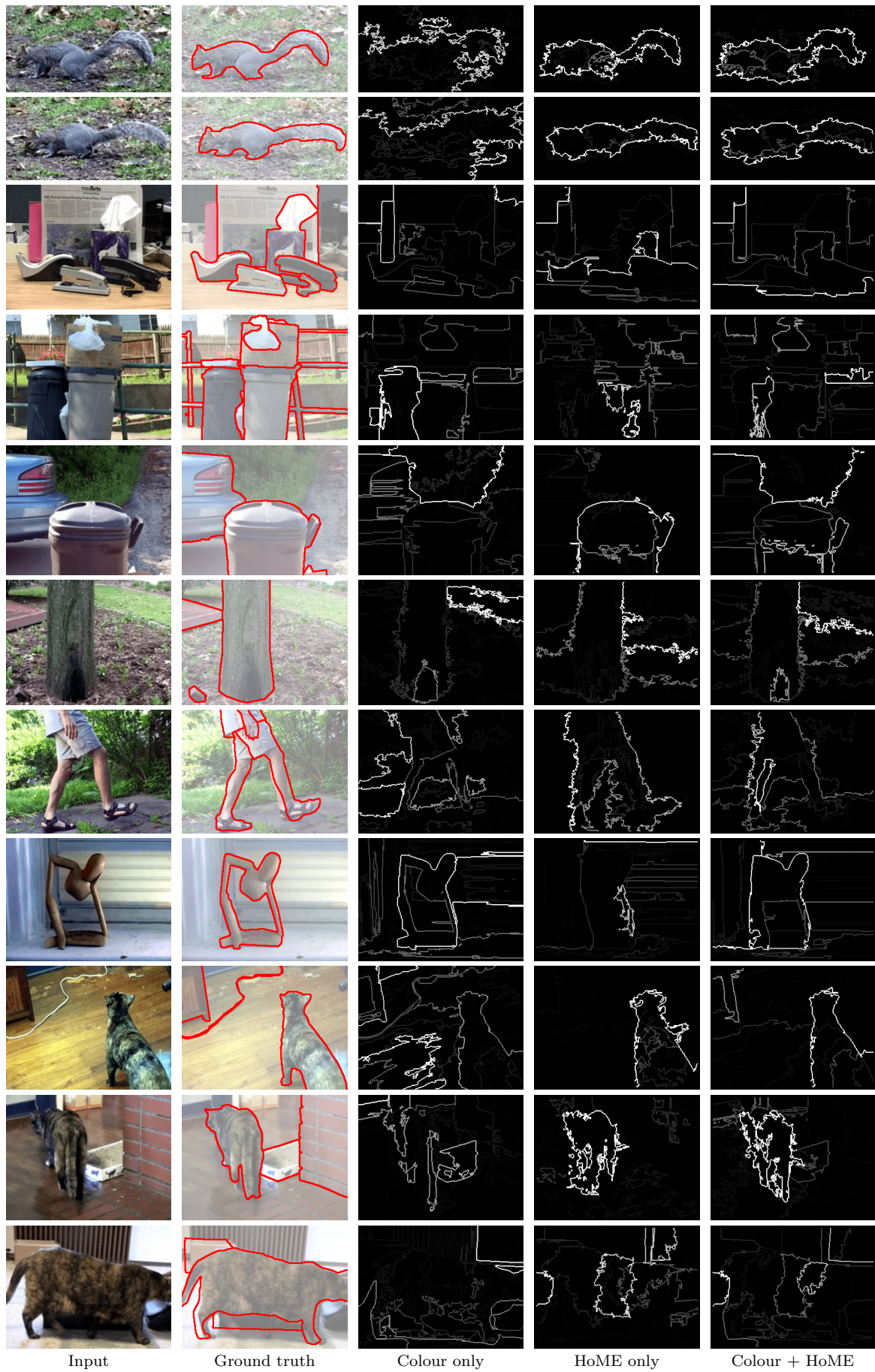


Figure 5: Results on the CMU dataset (continued).

References

- [1] A. Chan and N. Vasconcelos. Modeling, clustering, and segmenting video with mixtures of dynamic textures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(5):909–926, 2008.
- [2] A. B. Chan and N. Vasconcelos. Layered dynamic textures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(10):1862–1879, 2009.
- [3] J. Chen, G. Zhao, M. Salo, E. Rahtu, and M. Pietikäinen. Automatic dynamic texture segmentation using local descriptors and optical flow. *IEEE Trans. Image Processing*, 22(1):326–339, 2013.
- [4] X. He and A. Yuille. Occlusion boundary detection using pseudo-depth. In *ECCV*, pages 539–552. 2010.
- [5] G. Palou and P. Salembier. Depth ordering on image sequences using motion occlusions. In *ICIP*, pages 1217–1220. IEEE, 2012.
- [6] M. Sargin, L. Bertelli, B. Manjunath, and K. Rose. Probabilistic occlusion boundary detection on spatio-temporal lattices. In *ICCV*, pages 560–567, 2009.
- [7] A. N. Stein and M. Hebert. Occlusion boundaries from motion: Low-level detection and mid-level reasoning. *IJCV*, pages 325–357, 2009.
- [8] P. Sundberg, T. Brox, M. Maire, P. Arbelaez, and J. Malik. Occlusion boundary detection and figure/ground assignment from optical flow. In *CVPR*, pages 2233–2240, 2011.
- [9] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof. Anisotropic huber-L1 optical flow. In *BMVC*, pages 1–11, 2009.