Piecewise Rigid 3D Scene Flow

Christoph Vogel Konrad Schindler
Photogrammetry & Remote Sensing, ETH Zurich
{vogel,konrad.schindler}@geod.baug.ethz.ch

Stefan Roth
Department of Computer Science, TU Darmstadt
sroth@cs.tu-darmstadt.de

Abstract

Estimating dense 3D scene flow from stereo sequences remains a challenging task, despite much progress in both classical disparity and 2D optical flow estimation. We introduce a novel model that represents the dynamic 3D scene by a collection of planar, rigidly moving, local segments. We demonstrate the benefits of our model on different real-world image sets, including the challenging KITTI benchmark. We achieve leading performance levels, exceeding competing 3D scene flow methods, and attain better 2D motion estimates than all tested dedicated optical flow techniques.

1. Introduction

Scene flow estimation is defined as the extraction of dense 3D shape and 3D motion information from image data, acquired by two (or more) cameras at two (or more) time steps [7]. Applications include motion capture and analysis, driver assistance and autonomous navigation, and virtual or augmented reality. 3D scene flow can be regarded as a generalization of two classical computer vision problems, stereo and 2D optical flow estimation. In the light of recent developments in stereo [2, 11] and optical flow [*e.g.* 3, 8], the performance of 3D scene flow estimation [*e.g.*, 1, 10] still has been lagging behind. Despite the fact that more images are used, scene flow estimation from stereo sequences has not been able to outperform dedicated stereo and optical flow techniques at their respective tasks.

To address some of the limitations of existing techniques, we introduce a novel model that represents the dynamic 3D scene by a collection of planar, rigidly moving, local segments. Consequently our model implicitly represents a scene as regions with a consistent geometry and motion pattern, a property we assume to hold for most scenes of interest. Scene flow estimation can then be cast as jointly estimating a segmentation into planar and rigidly moving regions, and the 3D geometry and rigid motion parameters of these segments.

To the best of our knowledge, our method has been the first one to leverage the additional information contained in stereo sequences and outperform dedicated stereo and optical flow algorithms on realistic images.

Figure 1. Jointly estimated 3D geometry, 3D motion vectors, and superpixel boundaries, rendered from a different viewpoint.

2. Piecewise Rigid Model for 3D Scene Flow

In our model the scene is parameterized as a collection of piecewise planar regions, each moving rigidly over time. In order to fit the plane and motion parameters (9 unknowns per segment) reliably, a larger support is required, which is provided by an initial superpixel segmentation of the reference image. From that initial segmentation we create a large set of candidate *planes* in 3D object space, each with an associated rigid motion. Similar to fusion-based approaches to 2D optical flow [5], scene flow estimation is then cast as a labeling problem, which assigns each pixel to a segment and each segment to a rigidly moving 3D plane.

As is common in correspondence problems, we define an energy function including a data term to ensure a consistent appearance of the surface across all views, and a regularization term to encourage piecewise smooth geometry and motion. Additionally we add a boundary term assessing the quality of the segmentation into regions, and explicitly reason about occlusions both at the segment and pixel level. Starting from a set of superpixels, the energy is minimized in two steps: first a rigidly moving plane from a set of *proposals* is selected for each segment, keeping the pixel-to-segment assignment fixed; then, pixels are re-assigned

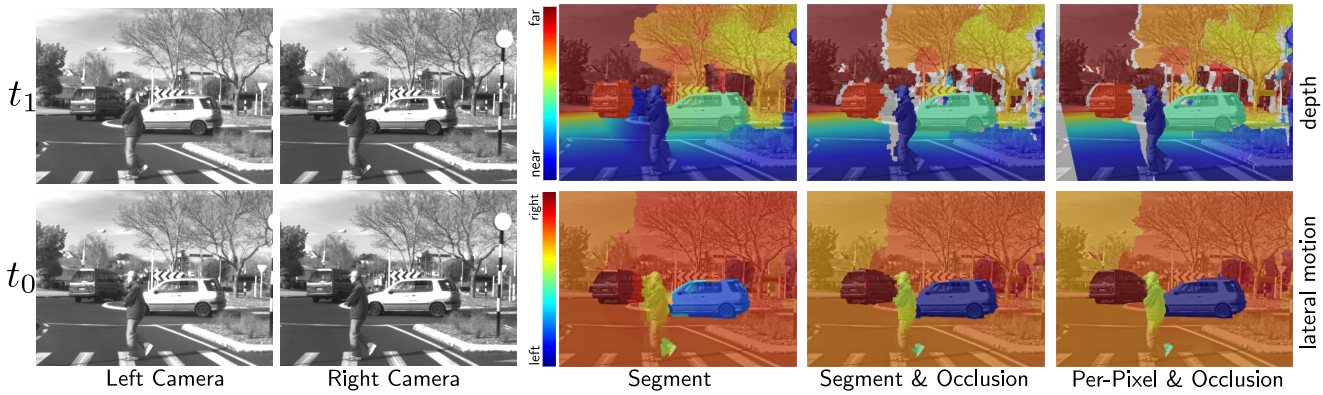

Figure 2. Example scene from [6]: Processing steps and final result of piecewise rigid scene flow estimation. Estimated depth and the lateral 3D motion component are shown. Occlusion areas are highlighted in white.

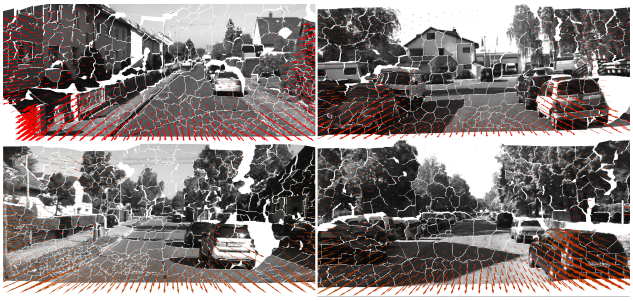

Figure 3. Example result from [4]: 3D reconstructions, rendered from a slightly different viewpoint.

to the best-fitting segment, while the planar geometry and motion of each segment is kept constant. Taken together the two inference steps lead to an over-segmentation of the scene which induces pixel correspondences that are consistent across views, and at the same time adheres to smoothness constraints.

3. Experiments

To visually demonstrate the performance of the proposed model on realistic data, we report qualitative results on a street scene from [6] and [4]. The pictures (Fig. 1, Fig. 3) show the estimated 3D scene flow, and the results after each individual stage of the inference (Fig. 2).

For quantitative evaluation we have tested the proposed approach on two-frame stereo pairs from the KITTI dataset [4] [1], where the method ranked 1st out of 30 published approaches for optical flow in all measures, and 3rd out of 31 published methods in stereo. Similar performance is only achieved by [12], which can only handle epipolar motion (whereas our method can cope with independent object motion, see Fig. 2).

[1] http://www.cvlibs.net/datasets/kitti/

4. Conclusion

We have presented a formulation of the 3D scene flow as a collection of rigidly moving, planar segments in 3D space, subject to smoothness constraints on both the geometry and the motion of the scene.

This work has been published at ICCV 2013 [9]. In the future we aim to extend the framework to more than two consecutive frames.

References

[1] T. Basha, Y. Moses, and N. Kiryati. Multi-view scene flow estimation: A view centered variational approach. In *CVPR'10*.

[2] M. Bleyer, C. Rother, P. Kohli, D. Scharstein, and S. N. Sinha. Object stereo – Joint stereo matching and object segmentation. In *CVPR 2011*.

[3] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV 2004*.

[4] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? In *CVPR 2012*.

[5] V. Lempitsky, S. Roth, and C. Rother. FusionFlow: Discrete-continuous optimization for optical flow estimation. In *CVPR 2008*.

[6] T. Vaudrey, C. Rabe, R. Klette, and J. Milburn. Differences between stereo and motion behaviour on synthetic and real-world stereo sequences. In *IVCNZ 2008*.

[7] S. Vedula, S. Baker, R. Collins, T. Kanade, and P. Rander. Three-dimensional scene flow. In *CVPR 1999*.

[8] C. Vogel, S. Roth, and K. Schindler. An evaluation of data costs for optical flow. In *GCPR 2013*.

[9] C. Vogel, S. Roth, and K. Schindler. Piecewise rigid scene flow. In *ICCV 2013*.

[10] C. Vogel, K. Schindler, and S. Roth. 3D scene flow estimation with a rigid motion prior. In *ICCV 2011*.

[11] K. Yamaguchi, T. Hazan, D. McAllester, and R. Urtasun. Continuous Markov random fields for robust stereo estimation. In *ECCV 2012*.

[12] K. Yamaguchi, D. McAllester, and R. Urtasun. Robust monocular epipolar flow estimation. In *CVPR 2013*.