

Dense Scene Flow from Stereo Disparity and Optical Flow

Extended Abstract

René Schuster
Oliver Wasenmüller
Didier Stricker

DFKI - German Research Center for Artificial Intelligence
firstname.lastname@dfki.de

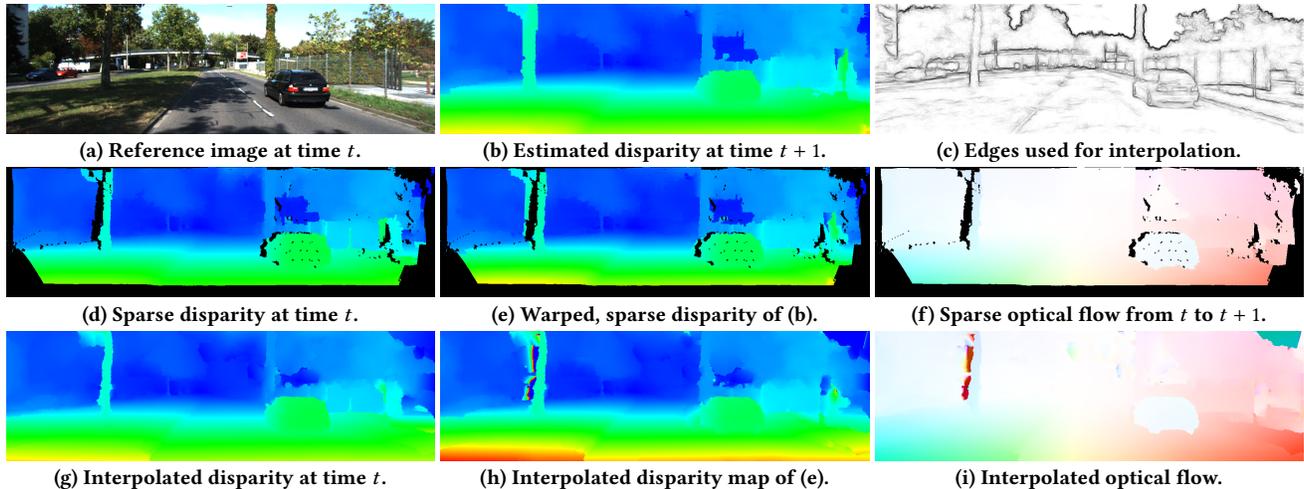


Figure 1: Optical flow and stereo disparity are combined to sparse scene flow by warping. Edge-aware interpolation is used to reconstruct a dense scene flow field.

ABSTRACT

Scene flow describes 3D motion in a 3D scene. It can either be modeled as a single task, or it can be reconstructed from the auxiliary tasks of stereo depth and optical flow estimation. While the second method can achieve real-time performance by using real-time auxiliary methods, it will typically produce non-dense results. In this representation of a basic combination approach for scene flow estimation, we will tackle the problem of non-density by interpolation.

1 INTRODUCTION

The problem of scene flow estimation in computer vision is the reconstruction of 3D geometry and 3D motion based on a sequence of stereo images (see Figure 2). While some papers argue that a joint estimation of geometry and motion will yield more consistent results, this paper estimates scene flow by the combination of stereo disparity and optical flow as it was already done in [12]. The sub-tasks are considered computationally less expensive and the combination itself is negligible in terms of run-time. Together with

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ACM Computer Science in Cars Symposium, 2018, Munich, Germany

© 2018 Copyright held by the owner/author(s).

a considerably fast and accurate interpolation it is even possible to reconstruct dense scene flow from stereo disparity and optical flow (see Figure 1).

This is an extension of the already published work in [12]. We will use the interpolation of SceneFlowFields (SFF) [14] to reconstruct dense scene flow from the combination of stereo disparity and optical flow. Relevant related work can be found in the original paper [12].

2 DENSE RE-COMBINATION

Sparse Combination. Since scene flow is the description of 3D geometry and 3D motion for every pixel of an image, it can be represented by two vector fields consisting of the 3D position and the 3D displacement in world space. An alternative representation can be given in image space by optical flow, disparity, and disparity change. Given the camera intrinsics and extrinsics, these two representations are equivalent and can be transformed into each other. Consequently, direct computation of optical flow and disparity solves two sub-tasks of scene flow estimation. What is missing is the change of disparity Δd_t^{t+1} that together with the disparity d_t yields the disparity at the next time step $d_t^{t+1} = d_t + \Delta d_t^{t+1}$ for each pixel of the reference time step. However, a disparity map d_{t+1} with reference to the next time step can be computed directly and together with optical flow which relates corresponding pixels

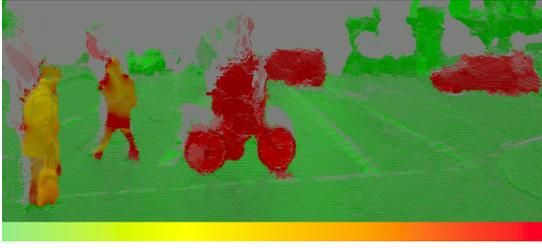


Figure 2: Point cloud visualization of dense scene flow result. Color indicates velocity (green: slow, red: fast).

Table 1: Comparison of different interpolation schemes. Percentage of outliers on KITTI training data.

Interpolation	D1	D2	Fl	SF	Density
<i>sparse</i>	4.4	8.3	9.0	12.7	84.23 %
<i>kitti</i>	11.0	17.5	19.9	23.8	100.0 %
<i>full</i>	6.5	12.6	16.3	19.6	100.0 %
<i>motion</i>	4.9	13.2	16.7	20.4	100.0 %
<i>disp-affine</i>	4.9	13.2	15.5	20.8	100.0 %
<i>disp-plane</i>	4.9	13.5	15.5	21.2	100.0 %

between both time steps, this disparity map can be warped to the reference frame:

$$d_t^{t+1}(x, y) = d_{t+1}(x + u, y + v). \quad (1)$$

$(u, v)^T$ are the optical flow components at pixel $(x, y)^T$. Thus, we can reconstruct scene flow from optical flow and two disparity maps. Bi-linear interpolation is used to warp disparity values from sub-pixel positions. The only problem is that the reconstruction will fail if the optical flow leaves the image boundaries or where the scene is occluded in the next time step. As a result, the reconstruction approach will produce a non-dense scene flow field (cf. Figures 1d to 1f).

Interpolation. The recently published SceneFlowFields [14] estimates scene flow by sparse-to-dense interpolation. We can utilize the edge-aware interpolation algorithm to reconstruct dense scene flow from the sparse combination of disparity and optical flow. SFF interpolates scene flow in two steps. First, local planes are estimated based on the known scene flow to interpolate geometry. Secondly, local affine transformations are estimated to describe the 3D motion.

Depending on the auxiliary methods that are used for optical flow and disparity estimation, the results for these tasks are already dense. Only the warped disparity map d_t^{t+1} is non-dense. That leaves several options for the interpolation which we all compare in Table 1:

- Using the default interpolation algorithm of the KITTI submission system (*kitti*).
- Interpolating all sub-tasks where d_t^{t+1} has gaps (*full*).
- Interpolating the 3D motion only (flow + disparity change) according to local affine 3D transformation models (*motion*).

Table 2: Results according to KITTI Scene Flow Benchmark [8] in percentage of outliers.

Method	D1	D2	Fl	SF	Run-time
ISF [2]	4.5	6.0	6.2	8.1	600 s
PRSM [18]	4.3	6.7	6.7	9.0	300 s
OSF+TC [10]	5.0	6.8	7.0	9.2	3000 s
OSF18 [9]	5.3	7.1	7.4	9.7	390 s
SSF [11]	4.4	7.0	7.1	10.1	300 s
OSF [8]	5.8	7.8	7.8	10.2	3000 s
FSF+MS [16]	6.7	9.9	11.3	15.0	2.7 s
CSF [7]	6.0	10.1	13.0	15.7	80 s
SFF [14]	6.6	10.7	12.9	15.8	65 s
PRSF [17]	6.2	12.7	13.8	16.4	150 s
Ours	6.6	14.4	16.6	20.7	36 s
SGM+SF [5, 6]	6.8	15.6	21.7	25.0	2700 s
PCOF+LDOF [4]	8.5	21.0	18.3	29.3	50 s
SGM+FF+ [12]	13.4	27.8	22.8	33.6	29 s
SGM+C+NL [5, 15]	6.8	28.25	35.6	40.3	270 s
SGM+LDOF [3, 5]	6.8	28.6	39.3	43.7	86 s

- Using the affine 3D transformations to interpolate the warped disparity only (*disp-affine*).
- Only interpolating the warped disparity map using a local plane model (*disp-plane*).

Results. Remarkable about the different concepts of interpolation is that the joint interpolation (an example result of this variant is given in Figures 1g to 1i) produces overall the best scene flow estimate, though the different sub-results are less accurate than for some other interpolation strategies (see Table 1). That supports the general paradigm that scene flow should be estimated jointly. Further, we want to highlight that the sparse combination results are very accurate. With the steady improvement of methods for the auxiliary tasks, scene flow estimation by re-combination gets better also. In this paper, we have used SPS-stereo [19] for disparity estimation and FlowFields++ [13] for the optical flow tasks. Both are ranked higher than the respective auxiliary methods (SGM [5] and FF+ [1]) that have been used in the original paper [12]. Due to that and because the interpolation algorithm we use in this paper is more sophisticated, our dense scene flow estimate from stereo disparity and optical flow is ranked higher in the official public KITTI Scene Flow Benchmark [8] than the original submission as shown in Table 2.

Run-time. The run-time of our approach consists of 29 s for FlowFields++ [13], 2 s for each disparity map computed with SPS-st [19], and 3 s for dense interpolation with SceneFlowFields [14].

3 CONCLUSION

Scene flow estimation by combination of stereo disparity and optical flow is as fast as the auxiliary methods are. The sparse and accurate results can be interpolated to a dense scene flow field with competitive performance. Improvements in stereo algorithms, optical flow estimation, and scene flow interpolation will directly improve the combination approach as presented in this work.

REFERENCES

- [1] Christian Bailer, Bertram Taetz, and Didier Stricker. 2017. Optical Flow Fields: Dense Correspondence Fields for Highly Accurate Large Displacement Optical Flow Estimation. *arXiv preprint arXiv:1703.02563* (2017).
- [2] Aseem Behl, Omid Hosseini Jafari, Siva Karthik Mustikovela, Hassan Abu Alhaja, Carsten Rother, and Andreas Geiger. 2017. Bounding Boxes, Segmentations and Object Coordinates: How Important is Recognition for 3D Scene Flow Estimation in Autonomous Driving Scenarios?. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [3] Thomas Brox and Jitendra Malik. 2011. Large displacement optical flow: Descriptor matching in variational motion estimation. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)* (2011).
- [4] Maxime Derome, Aurelien Plyer, Martial Sanfourche, and Guy Le Besnerais. 2016. A prediction-correction approach for real-time optical flow computation using stereo. In *German Conference on Pattern Recognition (GCPR)*.
- [5] Heiko Hirschmuller. 2008. Stereo processing by semiglobal matching and mutual information. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)* (2008).
- [6] Michael Hornacek, Andrew Fitzgibbon, and Carsten Rother. 2014. SphereFlow: 6 DoF scene flow from RGB-D pairs. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [7] Zhaoyang Lv, Chris Beall, Pablo F Alcantarilla, Fuxin Li, Zsolt Kira, and Frank Dellaert. 2016. A continuous optimization approach for efficient and accurate scene flow. In *European Conference on Computer Vision (ECCV)*.
- [8] Moritz Menze and Andreas Geiger. 2015. Object scene flow for autonomous vehicles. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [9] Moritz Menze, Christian Heipke, and Andreas Geiger. 2018. Object Scene Flow. *Journal of Photogrammetry and Remote Sensing (JPRS)* (2018).
- [10] Michal Neoral and Jan Šochman. 2017. Object Scene Flow with Temporal Consistency. In *Computer Vision Winter Workshop (CVWW)*.
- [11] Zhile Ren, Deqing Sun, Jan Kautz, and Erik B. Sudderth. 2017. Cascaded Scene Flow Prediction using Semantic Segmentation. In *International Conference on 3D Vision (3DV)*.
- [12] René Schuster, Christian Bailer, Oliver Wasenmüller, and Didier Stricker. 2018. Combining Stereo Disparity and Optical Flow for Basic Scene Flow. In *Commercial Vehicle Technology Symposium (CVT)*.
- [13] René Schuster, Christian Bailer, Oliver Wasenmüller, and Didier Stricker. 2018. FlowFields++: Accurate Optical Flow Correspondences Meet Robust Interpolation. In *International Conference on Image Processing (ICIP)*.
- [14] René Schuster, Oliver Wasenmüller, Georg Kuschik, Christian Bailer, and Didier Stricker. 2018. SceneFlowFields: Dense Interpolation of Sparse Scene Flow Correspondences. In *Winter Conference on Applications of Computer Vision (WACV)*.
- [15] Deqing Sun, Stefan Roth, and Michael J Black. 2014. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision (IJCV)* (2014).
- [16] Tatsunori Tani, Sudipta N. Sinha, and Yoichi Sato. 2017. Fast Multi-frame Stereo Scene Flow with Motion Segmentation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [17] Christoph Vogel, Konrad Schindler, and Stefan Roth. 2013. Piecewise rigid scene flow. In *International Conference on Computer Vision (ICCV)*.
- [18] Christoph Vogel, Konrad Schindler, and Stefan Roth. 2015. 3D Scene Flow Estimation with a Piecewise Rigid Scene Model. *International Journal of Computer Vision (IJCV)* (2015).
- [19] Koichiro Yamaguchi, David McAllester, and Raquel Urtasun. 2014. Efficient joint segmentation, occlusion labeling, stereo and flow estimation. In *European Conference on Computer Vision (ECCV)*.