Towards Flow Estimation in Automotive Scenarios

René Schuster[1], Oliver Wasenmüller[1], Georg Kuschk[2], Christian Bailer[1], Didier Stricker[1]

[1] DFKI – German Research Center for Artificial Intelligence
[2] BMW Group
firstname.lastname@{dfki,bmw}.de

Abstract. In this position paper, we discuss different motion estimation methods in general and especially for automotive scenarios. The advantage of 3D scene flow over 2D optical flow is explained by a typical use case. An indirect method for scene flow reconstruction from stereo disparity and optical flow is presented along with its main issue. Further, we describe how a direct estimation can benefit the overall scene flow result. These observations support our statement that research should focus more on direct 3D motion estimation in the future.

Keywords: Optical flow, Scene flow, Stereo

1 Introduction

Recent development in the area of Computer Vision has enabled a growing set of vision based applications for the automotive industry. Especially in the area of Advanced Driver Assistance Systems (ADAS), major innovations were released in recent years. They target topics of comfort and safety and close the gap towards autonomous driving considerably. All these systems have in common that they require a reliable and detailed perception and understanding of a vehicles environment.

Motion estimation of the traffic environment is one major aspect of this challenge. Typically, it is differentiated between optical flow, which estimates a 2D motion field, and scene flow, which operates in the 3D domain. While there has been steady evolution in motion estimation with respect to the 2D image space, there was comparably little advance in the estimation of motion in the 3D space [1]. Yet, there exist ADAS applications which would benefit substantially from that information.

In this position paper, we make the following two assertions. First, the additional information in scene flow opposed to optical flow is beneficial to ADAS. Secondly, the estimation of scene flow formulated as a single problem is superior to the recombination of separate solutions to depth estimation from stereo disparity and optical flow tasks. To this end, we describe the shortcomings of optical flow, illustrate the main issues of the reconstruction of scene flow from stereo and optical flow information, and list advantages of motion estimation through a full scene flow representation.

2 Motion Estimation

In this section, we describe the different variants of motion estimation along with their advantages and disadvantages. When referring to depth estimation or stereo disparity, we assume a stereo camera setup and the principles of stereo reconstruction as described in [2].

2.1 2D Optical Flow

Optical flow is the apparent 2D motion field between a scene and an observer. It associates an image point $\mathbf{x} = (x, y)$ with its shift in position $\mathbf{u} = (u, v)$ from time t to $t + 1$ (cf. Figure 1(a)).

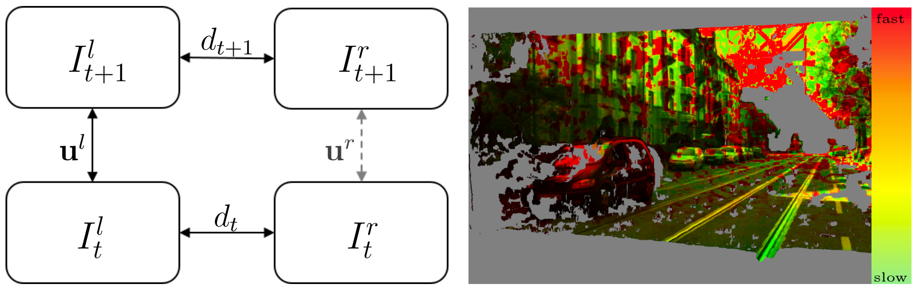
Since optical flow is only a representation of 2D motion from one pixel to another, it can not capture shifts in depth. This is a disadvantage because it can lead to misinterpretation of a traffic scene. The motion of a car approaching parallel to the viewing direction will not be detect by optical flow (except for a small zooming effect). Similar, depth information alone can be ambiguous. This motivates a motion representation in 3D space.

2.2 3D Scene Flow

Scene flow is often described as the extension of depth estimation over time and even more often as the generalization of optical flow to a third dimension [3]. An imaged point $\mathbf{x}$ of the 3D scene gets associated with a 3D vector $\mathbf{V} = (\Delta X, \Delta Y, \Delta Z)$ that represents the motion between two time steps relative to the camera, where – similar to optical flow – $\mathbf{x}$ is the projection of a 3D point $\mathbf{X} = (X, Y, Z) \in \mathbb{R}^3$ to image coordinates. In fact, optical flow is the projection of scene flow to an image plane.

The Naïve Approach. The missing 3D information to extend optical flow to scene flow can be obtained by depth estimation from stereo reconstruction. Given two disparity maps and the corresponding optical flow, it is possible to compute scene flow. These three inputs can be estimated separately from two stereo image pairs at consecutive time steps as illustrated in Figure 1(a). In this graphic, d_t and d_{t+1} are the disparity maps for time t and $t + 1$ respectively, and $\mathbf{u}^l$ is the optical flow for the left reference image. The change in disparity $\Delta d(x, y) = d_{t+1}(x + u, y + v) - d_t(x, y)$ and the 2D optical flow vectors can be projected to 3D space using the well-known pinhole camera model [2] to obtain a full 3D motion vector. Note that it is also possible to compute an optical flow field $\mathbf{u}^r$ for the right image, but that this is not necessary for the described naïve approach.

Figure 1(b) shows an exemplary scene flow result computed by the naïve approach, where each 3D point gets colored according to the magnitude of the associated 3D motion. The images are taken from the KITTI dataset [4,5] that is widely used for evaluation of different Computer Vision tasks in an automotive



(a) Image relations **(b)** Naïve scene flow

Fig. 1: (a) Relation of two stereo image pairs, (b) an exemplary result of scene flow reconstruction from stereo disparity and optical flow on KITTI dataset [4].

context. Depth is computed using the OpenCV implementation of Semi-Global Matching (SGM) [6] that has about 11 % outliers in the KITTI stereo metric, and optical flow is computed using FlowFields [7] which can be considered state-of-the-art.

Despite the comparable good input quality, the scene flow result (see Figure 1(b)) from combining stereo disparity and optical flow is noisy and unreliable. The reason for this is mainly that small errors in depth and optical flow sum up to a bigger error in the 3D flow. Assuming that disparity and optical flow are computed correctly up to a small error in the optical flow. This small error leads to an error in all three components of the scene flow, even in the depth shift, although the depth maps are correct. Expanding this thought experiment to noisy estimates for both disparity maps and the optical flow field clarifies why the naïve approach is extremely sensitive to smallest errors.

Direct Scene Flow Estimation. To overcome the issue of error accumulation, scene flow should be formulated as a single optimization problem such that consistency for all three components of the motion is ensured. Because scene flow is an under-determined problem, some regularization needs to be added. This is a popular practice in 2D optical flow estimation that will enforce smoothness over the flow field. This way, scene flow should in general yield better results than optimizing three different processes and combining them. In addition to the increased robustness, the direct solution is less sensitive to occlusions than the naïve approach. The recombination of scene flow is possible only if both disparity values – before and after motion displacement – are available. Close objects can cause large displacement between the left and right camera and thus lead to large occlusions, similar to fast objects in optical flow. This – by nature – leads to less dense scene flow when applying the naïve approach. The direct scene flow formulation can use all four images (cf. Figure 1(a)) to provide some mechanism for occlusion handling. Occlusions in one stereo image pair does not necessarily occur in the other and motion can be estimated from two view points.

There is an indirect empiric proof for the aforementioned advantages. A glance at the leader boards of the KITTI benchmark for scene flow shows that the naïve approach has been tried multiple times as combinations of SGM [6] with different methods for optical flow estimation [8,9,10] in a similar fashion we have done it. These approaches have been outperformed with a large margin by methods that estimate scene flow directly [3,11]. Considering the small number of recent publications about direct scene flow estimation, this is an even stronger evidence for our claim that these methods are superior to the naïve approach.

3 Conclusion

Because optical flow is not sufficient for some applications in the automotive industry, a motion representation in 3D space is beneficial. We have shown the main issue with scene flow estimation from depth and optical flow. Further, we have described how to overcome the issue and explained which other advantages arise from a direct solution to the scene flow task. The weaknesses of the naïve approach and the advantages of direct motion estimation in 3D lead us to the conclusion, that research should shift its focus from optical flow towards the scene flow problem so that a comparable amount of effort will be put into this field in the future.

References

1. Yan, Z., Xiang, X.: Scene flow estimation: A survey. arXiv:1612.02590 (2016)
2. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. (2003)
3. Vogel, C., Schindler, K., Roth, S.: 3D scene flow estimation with a piecewise rigid scene model. International Journal of Computer Vision (2015)
4. Menze, M., Geiger, A.: Object scene flow for autonomous vehicles. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. (2015)
5. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? The kitti vision benchmark suite. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. (2012)
6. Hirschmuller, H.: Stereo processing by semiglobal matching and mutual information. Transactions on Pattern Analysis and Machine Intelligence (2008)
7. Bailer, C., Taetz, B., Stricker, D.: Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation. In: Proceedings of the International Conference on Computer Vision. (2015)
8. Brox, T., Malik, J.: Large displacement optical flow: descriptor matching in variational motion estimation. Transactions on Pattern Analysis and Machine Intelligence (2011)
9. Sun, D., Roth, S., Black, M.J.: A quantitative analysis of current practices in optical flow estimation and the principles behind them. International Journal of Computer Vision (2014)
10. Hornacek, M., Fitzgibbon, A., Rother, C.: Sphereflow: 6 dof scene flow from rgb-d pairs. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. (2014)
11. Menze, M., Geiger, A.: Object scene flow for autonomous vehicles. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. (2015)