"What Is Optical Flow for?": Workshop Results and Summary

Fatma Güney[1,5], Laura Sevilla-Lara[2,5], Deqing Sun[3,5], and Jonas Wulff[4,5]

[1] Oxford University, UK
[2] Facebook Research
[3] NVIDIA
[4] Massachusetts Institute of Technology
[5] **Equal contribution and alphabetical order**

1 Introduction

Traditionally, computer vision problems have been classified into three levels: low (image to image), middle (image to features), and high (features to analysis) [11]. Some typical low-level vision problems include optical flow [7], stereo [10] and intrinsic image decomposition [1]. The solution to these problems would then be combined to solve higher level problems, such as action recognition and visual question answering. For example, optical flow has been used as an input to structure from motion, action recognition, and visual effects.

Two recent developments in optical flow affect this general paradigm for solving computer vision.

First, highly accurate optical flow can finally be learned [3, 8, 14, 19]. In most high-level computer vision problems the state-of-the-art methods have been based on deep learning for a few years. However, in optical flow it is only within the last year that the top performing methods [8, 19] are end-to-end trainable networks. This opens a new research question of how should we integrate these learnable optical flow modules into large systems to solve the computer vision problem.

Second, top performing methods are now very accurate in optical flow benchmarks. For example in KITTI, the state-of-the-art method achieves 92% accuracy, and in Sintel the average end-point-error is below 5 pixels, in images that are 1024 pixels wide (or 0.4%). This leads to a series of questions about the evaluation of flow. For example, are current methods accurate enough for higher level applications? In other words, is optical flow solved? If not, how should we design new benchmarks in the future to ensure they are most useful for higher level applications?

The goal of this workshop is to revisit the original plans of when and how to use optical flow for computer vision applications in light of these recent developments. We invite members of both researchers and practitioners of optical flow, to learn about recent progress, and to address these questions under the general topic of "What is optical flow for?".

2 Survey and speaker results

To collect input on the leading question, "What is optical flow for?", we followed a two-stage approach. First, we solicited feedback from the community through a survey, sent out to researchers in the field of optical flow and general video analysis in advance to the workshop. Second, we invited 9 speakers, selected for their long experience in research on optical flow as well as applications, to answer the same question in short talks during the workshop. This section briefly summarizes the result of both.

2.1 Survey

In this section, we present the questions and the outcome of our survey on optical flow. In the survey, we grouped our questions in three main categories: the participant background, the current status and the future of optical flow. We follow the same organization here, by first explaining our goal in each subsection and then sharing our findings. While preparing the survey, our goal was to maximize the amount of information we collect while minimizing the amount of time the participant spends on answering the questions. With this purpose in mind, for almost all questions, we provided a set of options by asking the participant to select all that apply and also including the option "other" for the participant to write down their own answer if none of the options fit.

Participant Background In the first part of the questionnaire, we asked participants about their background and their relation to optical flow. Our goal was to find out how informed the participants are about optical flow. The main questions included their current position, their rating of their knowledge on optical flow, and what computer vision problems they have worked on. We also asked more specific questions related to optical flow algorithms they use and optical flow benchmarks they check.

In total, 45 people took the survey, more than half of the participants are graduate students (63.6%) followed by postdoctoral research assistants or researchers (20.5%), professors (9.1%), and researchers in industry. More than half of the participants (55.6%) said that they have used and implemented optical flow, the remaining participants said that they have used optical flow (33.3%) or they at least know optical flow (11.1%). In summary, we can conclude that all of our participants had a research background and they were knowledgeable about optical flow.

We identified a set of computer vision tasks related to optical flow and shaped our questions around them throughout the survey. The experiences of our participants were roughly equally distributed over these tasks: action localization or recognition (37.8%), object tracking (31.1%), video object detection (22.2%), next-frame prediction (26.7%), video semantic segmentation (22.2%), and other video related tasks (35.6%).

Current Status of Optical Flow The goal of our questions in this part is two-fold. First, we wanted to find out what participants think about the usefulness of optical flow for other computer vision tasks. We asked at which level these tasks benefit from optical flow and if participants know any examples of specific optical flow algorithms being used for any of these tasks. Our second goal was to find out about the problems related to optical flow, in particular in evaluation. We asked which property of an optical flow algorithm they consider more important, speed or efficiency or both. Lastly, we included questions related to their experiences with optical flow algorithms, what kind of strategies they follow to choose the optical flow algorithm they use and then the most common problems they encounter while using an optical flow algorithm.

In terms of usefulness of optical flow algorithms for the computer vision tasks we identified, most of our participants think that these tasks either definitely benefit from optical flow or it could potentially benefit.An interesting finding is the relative ordering of these tasks in benefiting from optical flow. Next-frame prediction sticks out as the task which benefit the most from optical flow followed by action recognition or localization, object tracking, video semantic segmentation, and video object detection. As expected, most of the participants (79.1%) said that both speed and accuracy are important for optical flow. This finding is supported by the outcome of another related question: how to choose the optical flow algorithm to use. The most effective factors were identified as the state-of-the-art on optical flow benchmarks (78.6%) and speed or being able to run on a GPU (76.2%). Following the importance of accuracy and speed for optical flow, 66.7% of the participants stated that they need a more accurate flow algorithm and 61.9% a faster flow algorithm. The more specific problems have been identified as errors at large displacements (52.4%), bad results in occluded regions (42.9%), not enough structure in the flow field (33.3%), problems in motion boundaries (28.6%), and artifacts, spurious objects in the flow field (26.2%). A considerable amount of participants (23.8%) have identified optical flow as their main bottleneck.

The Future of Optical Flow In the last part, we asked our participants to speculate about the future of optical flow, starting with whether we need perfectly accurate optical flow. Next, we asked what we need to include in evaluation of optical flow that we currently do not consider, possibly more concerned about other tasks and the 3D world. Lastly, we asked what would eventually solve optical flow: better models, different learning strategies, or more data. We finished the survey by asking the participants to describe how they would use a *perfect* optical flow algorithm.

Almost half of the participants (48.8%) think that we need perfectly accurate optical flow only for certain applications such as computational photography or medical imaging. Only 17.1% of the participants think that current metrics are enough while the majority agrees that we need better means of evaluating optical flow. For this purpose, we suggested a set of options including performance on specific regions in the image (61%), robustness against noise (53.7%), per-

formance at different levels of motion blur (41.5%), performance as input to another task (36.6%), different types of camera motion (36.6%), performance on mostly motion oriented tasks (26.8%), the 3D structure of the world (26.8%), and performance with respect to adverse conditions (26.8%). The last question was about choosing more likely directions that could eventually solve optical flow. The highest percentage belongs to unsupervised or self-supervised learning with 62.8%, followed by better models and better representations for optical flow with 48.8% for each.

To summarize, our participants come from research background with experience in computer vision problems related to optical flow. Most of the participants agree that video-related computer vision tasks benefit or could potentially benefit from optical flow. The accuracy and the speed are identified as the two most important factors in choosing which optical flow algorithm to use as well as regarding the problems when using optical flow algorithms. Following that, recent deep learning methods e.g. FlowNet variants which are both fast and fairly accurate are frequently employed for various tasks, despite being new. Most of the participants think that we need better, more specific ways of evaluating optical flow.

2.2 Speakers

Both Michael Black and Jitendra Malik pointed out that our leading question, "What is optical flow for?", has been addressed in the past [12, 2]. In biology, research on motion perception and its purpose goes back at least as far as Gibson [5]; a list of what optical flow might be useful for in a biological system was given by Nakayama [12]. He identified 7 areas:

- Reasoning about the 3D structure of the environment
- Computing time to collision
- Image segmentation
- Computing ego-motion
- Computing saliency; control attention and eye movements
- Increasing contrast sensitivity
- Detect the motion of objects

As Jitendra Malik pointed out, this list, despite being over 30 years old, still applies today. Beyond this broad list, the speakers identified three large areas with close connection to optical flow: reasoning about the three-dimensional world, action understanding, and visualization and visual effects.

The persistency of the three-dimensional world As pointed out by Michael Black, an important distinction needs to be made between *optical flow* and the *motion field*. The first describes the motion of visual signals on the image plane, while the second models the motion of the three-dimensional scene relative to the observer, as projected into the two-dimensional image plane. With this distinction, a highlight moving across the surface of a static object would have induced

non-zero optical flow, although the motion field is zero. In practice, however, both terms are often used interchangeably, and optical flow often refers to the motion field. The motion field establishes persistence in the world and models which parts of the scene correspond across time. Thus, it can be used to extract multiple views of the same object, or to reason about foreground/background assignment at object boundaries.

Thomas Brox echoed this use of the motion field to obtain a 3D representation of the scene and presented a learned approach to 3D reconstruction, DeMoN [20]. An important input to DeMoN is the optical flow itself, which is computed before feeding it to the main reconstruction pipeline; this two-stage approach is advantageous since it allows separate pre-training of the optical flow computation network, which in itself is non-trivial. Another use for optical flow is as an auxiliary learning task [25]. In this setting, predicting optical flow in addition to the actual target task (in this case, a camera pose update) provides additional gradients to the network and hence makes the training more effective.

Going beyond the reconstruction of a static three-dimensional environment, Lourdes Agapito described how flow can help to distinguish moving objects from a static scene and reconstruct both. In this application, three-dimensional, deformable shapes are modelled using a low-dimensional set of deformation bases; the deformation and camera pose can then be fitted to the optical flow field, yielding a reconstruction of the full, non-rigid scene, and thus pointing towards a full semantic understanding of the scene using optical flow as input.

Action understanding One of the classical applications of optical flow is to classify and understand actions of people in videos. Actions and activities are inherently temporal processes, and the underlying assumption is that optical flow can be used to compute a motion signature specific to a particular action.

Cordelia Schmidt pointed out that using optical flow as an additional input indeed helps. Interestingly, however, the accuracy of the flow itself (in terms of EPE) does not have a large impact on the accuracy of the classification. It is therefore questionable whether heavy computation should be invested in computing even better optical flow in order to improve the results of action recognition. Consequently, she presented recent work on action recognition [17] that does not require pre-computation of optical flow, but uses only a stream of frames as input. This sentiment was echoed by Laura Sevilla-Lara, who presented results that show that, while optical flow helps action recognition, randomly shuffling a sequence of frames containing an action does not, in fact, degrade performance to chance. This indicates that long-term motion signatures might not be as important as assumed. Instead, she suggested that what is important about optical flow in the context of action recognition is localization as object boundaries, as well as overall motion of the human body [15].

Kristen Grauman pointed to a different benefit of using optical flow in action recognition, in that it can serve as a coarse measure of saliency. Flow can therefore direct attention of an algorithm towards even fine details in the frame, which would otherwise get lost in the image. This mechanism works even when

predicting optical flow from a single image; while just hallucinated, this flow can nevertheless improve action recognition by steering a network towards important image regions [4].

Overall, Jitendra Malik estimated action classification to lag about 10 years behind object detection, judging from the classification rates alone [6], and identified as the two main problems the long tail distribution of actions as well as current algorithms' inability to process long-range motion.

Visualization Going back to the distinction between the motion field (the projected motion of the 3D environment) and optical flow (the motion of the 2D visual stimulus on the image plane), Michael Black pointed out that the later is important for artistic and visual applications. As an example, if optical flow is used for temporal resampling in order adjust frame rates between different playback and recording devices, it is important to take the motion of the image into account: The motion of a highlight on a static sphere should be properly interpolated, too. Similarly, applying optical flow to warp and deform images opens up interesting creative possibilities in the temporal domain, such as creating the flowing color-like effect in the movie *What dreams may come* [23], or to synthesize and transfer facial deformation in *The Matrix Reloaded* [21].

However, as pointed out by Richard Szeliski, for many such applications in artistic domains optical flow is currently not good enough, both in terms of accuracy as well as in terms of representation. For example, since most current optical flow methods use only two input frames, temporal consistency of flow-based visual effects that would satisfy the human visual system is often hard to achieve, and requires painstaking manual labor. Another example is the treatment of non-lambertian surfaces containing effects such as highlights, reflections, or subsurface scattering. All these effects are not well modelled using current optical flow energy terms, and algorithms therefore fail in the presence of such surface properties. Lastly, it is critical to be able to model more than one motion at each location, both for transparent motions such as reflections as well as for partially occupied pixels at motion boundaries. Especially the appearance at object boundaries is critical for sufficient visual quality for professional applications; Richard Szeliski hence called for novel optical flow benchmarks including these challenging scenarios.

Bill Freeman described a system that deals with the particular case of transparent motion, which uses a two-layer model of the scene to allow the user to take photos through obstacles such as fences [24]. Furthermore, he pointed out that tiny motions contain a lot of information about the physical properties of the world, such as oscillations of large structures and subtle change in appearance due to blood flow in a face. Properly magnified [22, 13], these subtle motions can be made visible, and hence open up new applications in structural analysis and healthcare. Interestingly, while motion is crucial for this task, it is never represented directly as optical flow, but instead encoded using hand-crafted [22] or learned [13] spatio-temporal filters. This squares with the suggestion from

the survey (see above) to explore different, novel representations of the motion beyond optical flow.

Other remarks Beyond these three main areas, several speakers mentioned other applications for optical flow. Jitendra Malik hypothesized that, to "solve vision", his bet is on unsupervised learning of motion and subsequently using motion as a supervisory signal for to learn other tasks. Kristen Grauman echoed this, and showed how motion can be used to improve training a per-pixel objectness classifier [9]. She also described 360° video compression as an additional application for flow; here, the video is stored as if projected onto the six sides of a cube, and optical flow is useful to determine the orientation of the cube to ensure best compressibility [16]. Lourdes Agapito talked about her working experiences with robotics companies and remarked that robotics is a sober exercise because the algorithms have to work in real scenarios. Robustness to real-life distortion is critical for the deployment of optical flow, yet is currently missing from the datasets.

Deqing Sun talked about an empirical study of CNN for optical flow, which shows that models matter, so does training [18]. The FlowNetC model, re-trained using the procedure of PWC-Net, outperforms the published FlowNet2 on Sintel final pass, altough FlowNetC is a sub-network of the much larger FlowNet2 model. He also discussed about recent changes to the training procedures of PWC-Net, which brings about 10 to 20 % improvement on Sintel and KITTI.

3 Panel Discussion

One audience asked about Unit-tests for optical flow. The discussion was leaning toward task-oriented metrics for specific applications. During the panel discussion, Bill Freeman posed an interesting challenge: "take a 15-second video, re-render it under different lighting conditions/viewpoints." A successful solution would require accurate reconstruction of the scene geometry, lighting, material properties, and motion. Richard Szeliski commented that, while standard frame-rate videos have become standard, extremely high-frame rate cameras may significant benefit specific applications, such as autonomous driving. The high frame rate would make many vision problems less challenging, such as motion and tracking.

4 Conclusion

In summary, the workshop was a reminder of wide variety of applications of optical flow: segmentation, action classification, visualization, medical imaging, depth estimation, just to name a few - and all of these applications have different requirements regarding accuracy and fidelity of the representation of motion.

Optical flow is not solved, not only in terms of the error in current benchmarks but in terms of impact for applications. Some applications like action

recognition may not benefit directly from better optical flow, but it is not clear if this is intrinsic to the problem, or a consequence of the choice of categories or current recognition networks. At the same time, other applications of flow do benefit from better flow, like visual effects, or non-rigid structure from motion.

Improving these applications may require new benchmarks and evaluation metrics that are application specific, and that give insight into the impact of optical flow progress for different applications. In addition to new benchmarks, it will be interesting to explore better representations of motion, beyond simple optical flow, that may be more fit to applications.

References

1. Barrow, H., Tenenbaum, J., Hanson, A., Riseman, E.: Recovering intrinsic scene characteristics. Comput. Vis. Syst **2**, 3–26 (1978)
2. Black, M.J.: Robust incremental optical flow. Ph.D. thesis, PhD thesis, Yale university (1992)
3. Dosovitskiy, A., Fischery, P., Ilg, E., Hazirbas, C., Golkov, V., van der Smagt, P., Cremers, D., Brox, T., et al.: FlowNet: Learning optical flow with convolutional networks (2015)
4. Gao, R., Xiong, B., Grauman, K.: Im2flow: Motion hallucination from static images for action recognition. In: CVPR (2018)
5. Gibson, J.J.: The perception of the visual world. (1950)
6. Gu, C., Sun, C., Ross, D.A., Vondrick, C., Pantofaru, C., Li, Y., Vijayanarasimhan, S., Toderici, G., Ricco, S., Sukthankar, R., et al.: Ava: A video dataset of spatio-temporally localized atomic visual actions. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2018)
7. Horn, B., Schunck, B.: Determining optical flow. Artificial Intelligence (1981)
8. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: FlowNet 2.0: Evolution of optical flow estimation with deep networks (2017)
9. Jain, S., Xiong, B., Grauman, K.: Pixel objectness. arXiv preprint arXiv:1701.05349 (2017)
10. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. pp. 674–679 (1981)
11. Marr, D.: Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. W.H. Freeman, New York, NY, USA (1982)
12. Nakayama, K.: Biological image motion processing: a review. Vision research **25**(5), 625–660 (1985)
13. Oh, T.H., Jaroensri, R., Kim, C., Elgharib, M., Durand, F., Freeman, W.T., Matusik, W.: Learning-based video motion magnification. In: The European Conference on Computer Vision (ECCV) (September 2018)
14. Ranjan, A., Black, M.J.: Optical flow estimation using a spatial pyramid network (2017)
15. Sevilla-Lara, L., Liao, Y., Guney, F., Jampani, V., Geiger, A., Black, M.J.: On the integration of optical flow and action recognition. arXiv preprint arXiv:1712.08416 (2017)
16. Su, Y.C., Grauman, K.: Learning compressible 360deg video isomers. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (June 2018)

17. Sun, C., Shrivastava, A., Vondrick, C., Murphy, K., Sukthankar, R., Schmid, C.: Actor-centric relation network. In: The European Conference on Computer Vision (ECCV) (September 2018)
18. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: Models matter, so does training: An empirical study of cnns for optical flow estimation. arXiv preprint arXiv:1809.05571 (2018)
19. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume (2018)
20. Ummenhofer, B., Zhou, H., Uhrig, J., Mayer, N., Ilg, E., Dosovitskiy, A., Brox, T.: Demon: Depth and motion network for learning monocular stereo. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017), http://lmb.informatik.uni-freiburg.de//Publications/2017/UZUMIDB17
21. Wachowski, L., Wachowski, L.: The Matrix Reloaded (2003)
22. Wadhwa, N., Rubinstein, M., Durand, F., Freeman, W.T.: Phase-based video motion processing. ACM Trans. Graph. (Proceedings SIGGRAPH 2013) **32**(4) (2013)
23. Ward, V.: What Dreams May Come (1998)
24. Xue, T., Rubinstein, M., Liu, C., Freeman, W.T.: A computational approach for obstruction-free photography. ACM Transactions on Graphics (Proc. SIGGRAPH) **34**(4) (2015)
25. Zhou, H., Ummenhofer, B., Brox, T.: Deeptam: Deep tracking and mapping. In: European Conference on Computer Vision (ECCV) (2018), http://lmb.informatik.uni-freiburg.de/Publications/2018/ZUB18