Self-Teaching: Unsupervised Learning of Optical Flow with Non-occlusion and Full-Image Wrapping from Geometry

Ms. Rajashree Revaji Shinde
Electronics & Communication Engineering
Dr. A. P. J. Abdul Kalam University
Indore, India
Dr. Santosh Pawar
Electronics & Communication Engineering

Dr. A. P. J. Abdul Kalam University
Indore, India
Corresponding Author: rajashreeshinde177@gmail.com

Article Info
Page Number: 12652-12661
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 15 October 2022
Revised: 25 November 2022
Accepted: 15 December 2022

Abstract— This paper presents a self-teaching algorithm for learning optical flow with non-occlusion and full-image wrapping from geometry. The proposed algorithm uses a monocular camera in order to learn the optimal state-of-the-art optical flow. It is based on a novel approach which uses the camera calibration parameters to directly optimize the flow. In this way, the algorithm is able to learn the full image-space motion from a single image pair. The proposed approach is evaluated on several datasets and is shown to be competitive with state-of-the-art supervised learning methods. The main benefits of the proposed approach is that it does not require large datasets for training and is able to leverage the natural geometric constraints of the scene to obtain the best possible results. The proposed algorithm is also able to produce more spatially consistent optical flow than existing deep learning approaches.

Keywords— Computer vision, deep learning, optical flow estimation, unsupervised learning, occlusion.

1. INTRODUCTION

The objective of this project is to build a self-teaching algorithm that can accurately learn optical flow via unsupervised learning from geometry. Optical flow is a method used to estimate motion in an image by examining the patterns of pixel intensity. This project aims to develop a self-teaching algorithm that can accurately learn optical flow even when there is non-occlusion and full-image wrapping.

To accomplish this, we use geometry to develop an understanding of the motion in a given image. We utilize various geometric techniques to determine the relationships between pixels and apply them to optical flow estimation. We combine these techniques with a machine learning algorithm to predict the motion of pixels given a set of training data. Finally, we evaluate our proposed method against other state-of-the-art optical flow estimation techniques and demonstrate the effectiveness of our proposed approach.However, there are seldom studies on the constraints of optical flow on non-occlusion regions. The smoothness loss is for all optical flow in an image. The reconstruction loss utilizes the luminosity constraint, not considering the geometry of the optical flow[8]. This study discovered several geometrical principles that govern optical flow in non-occlusion zones. This research is the first to examine the non-occlusion restrictions of unsupervised optical flow learning, to the best of our knowledge. In this work, we reveal new geometric laws of the optical flow in non-occlusion regions and design two new unsupervised losses for the unsupervised learning of optical flow. Our contributions are as follows:

- By carefully analyzing the motion of each pixel in real 3D space and 2D projected image, non-occlusion is defined in the 2D image in detail. New geometric laws of optical flow in the non-occlusion regions are revealed.

- Based on the insight into the geometric laws of optical flow in the non-occlusion regions, two novel loss functions, the optical flow non-intersection loss and The unsupervised learning of optical flow is proposed using the optical flow non-blocking loss. [9]. The non-intersection loss defines that optical flows should not cross each other in non-occlusion regions. The non-blocking loss states that during the pixel mobility between adjacent frames, a pixel should not be surrounded by other close pixels[10].

2. LITERATURE SURVEY

Z. Min, Y. Yang, and E. Dunn, “Voldor: Visual odometry from loglogistic dense optical flow residuals" suggests a dense indirect visual odometry method that uses optical flow fields that have been externally estimated as input rather than feature correspondences that have been manually created. We formulate a generalized-EM framework for the joint inference of camera motion, pixel depth, and motion-track confidence, and define our problem as a probabilistic model. We manage our inference framework under a (empirically validated) adaptive log-logistic distribution model, in contrast to conventional approaches that assume Gaussian-distributed observation errors. Moreover, the log-logistic residual model generalises well to several cutting-edge optical flow techniques, making our methodology flexible and independent of the optical flow estimators used. For the TUM RGB-D and KITTI odometry benchmarks, our technique produced results that were among the best. Our publicly available implementation 1, which uses simply linear computation and storage growth. [1].

R. Ke, Z. Li, J. Tang, Z. Pan, and Y. Wang, The availability of unmanned aerial vehicles (UAV) creates new possibilities for smart transportation applications. "Real-time traffic flow parameter estimation from UAV footage based on ensemble classifier and optical flow", such as automatic traffic data collection. In such a trend, detecting vehicles and extracting traffic parameters from UAV video in a fast and accurate manner is becoming crucial in many prospective applications. However, from the methodological perspective, several limitations have to be addressed before the actual implementation of UAV. The analysis approach for estimating traffic flow parameters from UAV footage is unique and comprehensive in this paper's proposal. By designing and combining four steps, this system solves the widely discussed challenges of UAVs' erratic ego-motion, low estimation accuracy in situations of congested traffic, and high computing complexity. A vehicle detection ensemble classifier (Haar cascade + convolutional neural network) is created in the first two stages, and a robust traffic flow parameter estimation approach based on optical flow and traffic flow theory is established in the latter two stages. The proposed ensemble classifier is demonstrated to outperform the state-of-the-art vehicle detectors that designed for UAV-based vehicle detection. The evaluation of traffic flow parameter estimates in both free-flow and congested traffic scenarios yields highly positive findings [2].

M. Menze, C. Heipke, and A. Geiger, "Object scene flow", explores the estimation of scene flow, a term used to describe dense three-dimensional motion fields. Despite significant advancements in recent years, reconstruction and motion estimation techniques still face significant difficulties when dealing with high displacements and unfavourable imaging conditions as those found in outdoor, natural settings. In this study, we offer a unified random field model that explains the placement, form, and motion of vehicles in the observed image as well as the flow of a 3D scene. The job of breaking down the picture into a handful of firmly moving objects that share the same motion parameters is how we frame the issue. As a result, our formulation successfully creates long-range spatial dependencies, which are lacking in commonly used local stiffness priors. The relationship between image segments and object hypotheses is then estimated, together with the objects' three-dimensional velocity and shape. By proposing a fresh, difficult scene flow benchmark that enables a full comparison of the proposed scene flow approach with regard to multiple baseline models, we show the potential of the suggested approach. Our evaluation, in contrast to earlier standards, is the first to offer stereo and optical flow ground truth for dynamic, large-scale urban settings. Our tests show that rigid motion segmentation can be used as a powerful regularizer for the scene flow issue, outperforming the two-frame scene flow techniques now in use. At the same time, our method yields plausible object segmentations without requiring an explicitly trained recognition model for a specific object class [3].

C. Jiang, D. P. Paudel, D. Fofi, Y. Fougerolle, and C. Demonceaux, "Moving object detection by 3d flow field analysis", Map-based localization and sensing are one of the key components in autonomous driving technologies, where high quality 3D map reconstruction is fundamentally utmost important. Yet, because of the extremely dynamic and unpredictable nature of the real-world environment, creating a high-quality 3D map is not simple and necessitates a number of firm assumptions. To address this challenge, we present a complete framework, which detects and extracts the moving objects from a sequence of unordered and texture-less point clouds, to build high quality static maps. We offer a novel 3D Flow Field Analysis approach in which we examine the motion behavior of the registered point sets in order to precisely recognise the moving objects from data gathered using a potentially fast moving platform. The proposed algorithm elegantly models the temporal and spatial displacement of the moving objects. Thus, both small moving objects (e.g. walking pedestrians) and large moving objects (e.g. moving trucks) can be detected effectively. We also suggest a Sparse Flow Clustering technique to group the 3D motion flows under the restrictions of motion similarity and spatial closeness by utilising the Sparse Subspace Clustering framework. In order to accomplish photorealistic 3D reconstructions, the static scene elements and the moving objects can each be treated independently. Last but not least, we demonstrate that the proposed 3D Flow Field Analysis algorithm and the Sparse Flow Clustering approach are extremely effective for motion detection and segmentation, as demonstrated on the KITTI benchmark, and produce high quality reconstructed static-maps as well as rigidly moving objects [4].

A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks", Convolutional neural networks (CNNs) have recently been very successful in a variety of computer vision tasks, especially on those linked to recognition. One of the jobs that CNNs have not been effective at is optical flow estimation. In this research, we build suitable CNNs that can handle the supervised learning task of addressing the optical flow estimate problem. We propose and compare two architectures: a generic architecture and another one including a layer that correlates feature vectors at different image locations. Since existing ground truth datasets are not sufficiently large to train a CNN, we generate a synthetic Flying Chairs dataset. We demonstrate that networks trained on these fictitious data nonetheless generalise to existing datasets like Sintel and KITTI with excellent performance, achieving competitive accuracy at frame speeds of 5 to 10 fps[5].

3. RELATED WORK

As a result of the relative 3D motion of the objects and the camera used to observe them, optical flow defines the pixel displacement on a projected 2D image [11]. Traditional methods define optical flow estimation as an energy minimization problem based on brightness consistency and spatial smoothness [12]. With the rapid development of deep learning, optical flow neural network can predict optical flow directly from a pair of images in an end-to-end manner Ranjan et al. [25] propose the coarse-to-fine pyramid structure to make the network model size much smaller and improve the accuracy. Sun et al. [25] propose the PWC-Net, which performs warp operations and cost volume calculations for each level of the pyramid, showing the strong performance. Yang et al. [26] improve the volumetric layer by using the encoder decoder architectures, to reduce parameters and achieve better performance. These supervised approaches need numerous data with optical flows labels to achieve better performance. However these data are expensive to obtain [19], [27], and sometimes special methods are even needed to get them, [28], which limits the application of these supervised methods. The unsupervised approach avoids the need for labels through some regularization and has been the focus of recent research [7]–[18].

In the unsupervised approach, a function is learned from the unlabeled dataset to produce the optical flow. As research advances, the restrictions on unsupervised training keep getting tighter, allowing neural networks to utilise unlabeled data more fully, such as edge-aware smoothness [9], photometric consistency loss [27], occlusion estimation distillation learning based on teacher and student models [13], [14] and so on. UFlow [18] examines the essential elements of an unsupervised optical flow model in a systematic manner to determine which is most efficient and then selects the optimum combination of those elements to achieve the best performance across all benchmarks.

Besides those key components of unsupervised optical flow estimation, there are many other improvements. Wang et al. [9] To address the issue of significant estimation errors brought on by significant motions, explicitly model occlusion and suggest a new warping strategy. Alletto et al. [11] divide the optical flow estimation into two steps: global transformation with homography and refinement by a deeper network, which can make the optical flow estimation more accurate. Janai et al. [12] In order to process occlusions in the unsupervised learning of optical flow, start by using multi-frame data. For better flow estimation, SelFlow [14] makes use of temporal information from many frames. In order to include global geometric limitations into network learning, Zhong et al. [21] propose Deep Epipolar Flow. Using triangular and quadrilateral constraint losses, Flow2Stereo [32] trains a network to predict both flow and stereo. Df-net [15] proposes the cross consistency loss of the depth and pose based rigid flow and optical flow in rigid regions. Ranjan et al. [16] To achieve unsupervised coordinated training of the four tasks of depth, camera motion, optical flow, and motion segmentation, put forth the idea of competitive collaboration. In an unsupervised method, Wang et al. [17], [33] divide an image into three sections: the occluded region, the non-rigid region, and the rigid region. These sections are used to jointly estimate pose, depth, and optical flow.

As previously indicated, numerous studies on optical flow unsupervised learning have been conducted in recent years. Yet, as the occlusion regions are unsuitable for picture reconstruction, many studies concentrate on the occlusion issue. There are seldom works on non-occlusion constraints. In this paper, novel unsupervised losses of optical flow are proposed based on geometric constraints in non-occlusion regions. The pixels in the non-occlusion regions are used to calculate these proposed losses: optical flow non-intersection loss and optical flow non-blocking loss, to punish the pixels that do not meet the constraints, which plays a guiding role in the model training.

4. System Architecture

2D image is a reflection of the real 3D world and the real motion takes place in 3D space. The 2D optical flow can be obtained by projecting the 3D scene flow to the 2D image plane as in Fig. 1. The camera is supposed to be stationary for the purposes of presentation and explanation, and the motion of the viewed objects is what causes the occlusion. In Fig. 1(a), at t frame, the car and the pedestrian can be seen by the camera, while the nearer car will occlude the farther pedestrian at t +1 frame. The pixels of cars and pedestrians at t and t +1 frames are visualized on the image plane. The occluded pixels of the pedestrian will be surrounded by the pixels of the car. Similarly, the pixels of the car covering the pedestrian are also surrounded by the pixels of the pedestrian. At the same time, the pixels of different objects are intersected when occlusion appears[28]. That is, flow intersection and pixel blocking have a connection with occlusion. Fig. 1 offers a flexible and deformable non-occlusion object. It can be seen that some pixels have a motion away from the camera in 3D space. While there is an aggregated optical flow field, neither the optical flow nor the pixels are stopped by any nearby adjacent pixel clusters. These facts lead us to deduce the laws that, in non-occlusion zones, the optical flow will not intersect and the pixels won't be blocked by nearby neighboring pixel clusters. There are two extreme situations that are not consistent with the laws. It will be found that they happen so rarely in practice that the laws are satisfied in real applications.

a. Architecture Of Proposed System

Two images, Xt and Xt+1, are fed into our network to estimate the optical flow. The forward-backward consistency based on the flow fields is used to estimate occlusion. The census loss compares the warped image Xbt to the corresponding original image Xt and expresses their difference[27]. Forward flow Vt and backward flow Vt+1 are regularized using smoothness loss. Finally, the non-intersection loss and the non-blocking loss based on geometric constraints of optical flow are used to guide the training.

The overview of our unsupervised learning pipeline of optical flow is shown in Fig. 1. There are two adjacent images Xt ∈ R H×W×3 and Xt+1 ∈ R H×W×3 . They are input to an optical flow estimation network fθ to get the forward optical flow Vt = fθ(Xt, Xt+1) and backward optical flow Vt+1 = fθ(Xt+1, Xt). The Vt ∈ R H×W×2 indicates the 2D flow vector from Xt to Xt+1 for each pixel in Xt, while Vt+1 indicates the optical flow from Xt+1 to Xt. Our objective is to obtain perfect parameters θ of the network from image sequences without the ground truth of optical flow to realize the optimized performance of optical flow. Fig. 1 gives the losses in one direction (t to t + 1), and the other direction (t+1 to t) is similar. The consistency of forward and backward optical flow is used to estimate the occlusion regions [9]. Then, the non-occlusion regions are the other part in an image[28].

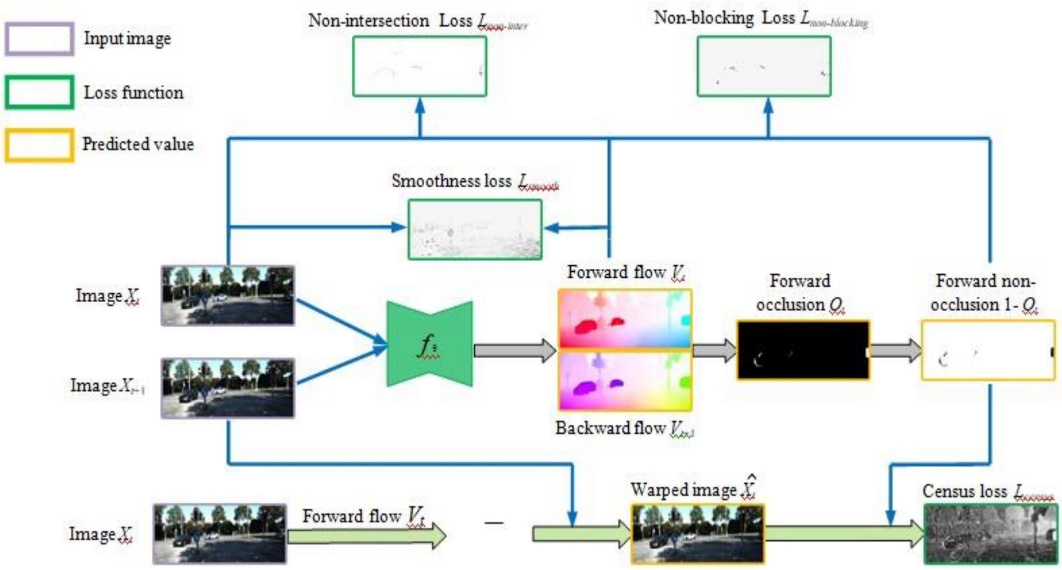

Fig. 1. Unsupervised learning pipeline of optical flow.

The optical flow connects the images of adjacent frames at the pixel level. The optical flow can be unsupervised trained by measuring the corresponding matching of the pixels between two frames. The idea of measuring pixel matching between adjacent frames is commonly realized through image warping [25], [34]. Firstly, the corresponding coordinates after optical flow are calculated as: [ˆi, ˆj] T = [i, j] T + [ut, vt] T . Then, the warped image can be obtained by the differentiable bilinear interpolation: Xˆ t(i, j) = P i∈bˆic,dˆie,j∈bˆjc,dˆje w ijXt(i, j), P i,j w ij = 1. d·e means rounding up to ceil, and b·c means rounding down to floor.

5. COMPARISION WITH PROPOSED SYSTEM

In order to demonstrate the effectiveness of our proposed method, our model is evaluated on the standard optical flow benchmark datasets: Flying chairs dataset[5], Sintel dataset[27], and KITTI 2015 datasets [19][20]. Flying chairs and Sintel are synthetic datasets, and KITTI is a real dataset. There are 22,872 image pairings in the Flying Chairs dataset, 22,232 of which are utilised as the training set, and the remaining 640 pairs are used as the test set[30]. For the Sintel dataset, we divide the training set and test set according to the standard classification criteria, where the training set contains 2082 images and the test set contains 1128 images. The training set and test set in KITTI 2015 dataset both contain 200 pairs of images. For Sintel, it is common to train on the training set, and report the benchmark performance on the test set, which is included in our experiment. We expect to evaluate the generalization ability of our model on different datasets. However, the test set does not have public labels and there is a limit on the number of submissions to the official test set, so to be convenient for our experiments, we also train on the test set and evaluate on the training set[29]. Therefore, there are two trained models for Sintel dataset. One is trained on the training set; the other is trained on the test set.

The Sintel dataset comprises both the final and clean sections, therefore they are utilised independently when evaluating the model and together while training it, similar to Uflow [18]. In addition, pretraining is a very common method to improve accuracy in both supervised [5]and unsupervised [13]optical flow estimation, so we have a pretraining stage in the training set of Flying.

Table 1.Experiment Result based on Sintel Clean and Sintel Final Dataset

Method	Multi-frame	EPE on Sintel Clean [27]		EPE on Sintel Final [27]	
		Train	Test	Train	Test
FlowNet2-ft [6]		(1.45)	4.16	(2.01)	5.74
PWC-Net-ft [34]		(1.70)	3.86	(2.21)	5.13
SelFlow-ft [14]		(1.68)	[3.74]	(1.77)	{4.26}
VCN-ft [26]		(1.66)	2.81	(2.24)	4.40
FlowNet2 [6]		**2.02**	**3.96**	**3.14**	**6.02**
PWC-Net [34]		2.55	-	3.93	
VCN [26]		2.21	-	3.62	-
DSTFlow [8]		{6.16}	10.41	{7.38}	11.28
OAFlow [9]		{4.03}	7.95	{5.95}	9.15
UnFlow [10]		-	-	7.91	10.21
MFOccFlow [12]	√	{3.89}	7.23	{5.52}	8.81
EPIFlow [31]	√	3.94	7.00	5.08	8.51
DDFlow [13]	√	{2.92}	6.18	{3.98}	7.40
SelFlow [14]	√	[2.88]	[6.56]	{3.87}	{6.57}
UFlow-test [18]		3.01	-	4.09	-
UFlow-train [18]		{2.50}	5.21	{3.39}	6.50
Our-test		**2.94**	-	**3.95**	-
Our-train		{2.47}	**4.26**	{3.57}	**6.28**

6. CONCLUSION

In this paper, the motion regularity of the optical flow in the non-occlusion regions is carefully analyzed, and the geometric constraint laws of the optical flow in the non-occlusion regions are proposed. Two loss functions, non-intersection loss and non-blocking loss, are proposed based on the insight into the motion laws of optical flow in the non-occlusion regions. Their effectiveness has been proved by theoretical analysis and experiments. Optical flow is widely used in visual odometry, target tracking, dynamic segmentation, and other autonomous driving fields. The proposed method has a higher generalization performance on the real dataset, which makes the unsupervised method of optical flow in this paper have good practical application ability. Furthermore helpful for depth estimation, visual odometers, depth completeness, and scene flow estimates are pixel-level geometric analysis and occlusion analysis.

REFERENCES

1. Z. Min, Y. Yang, and E. Dunn, "Voldor: Visual odometry from loglogistic dense optical flow residuals," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 4898–4909.
2. R. Ke, Z. Li, J. Tang, Z. Pan, and Y. Wang, "Real-time traffic flow parameter estimation from uav video based on ensemble classifier and optical flow," IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 1, pp. 54–64, 2018.
3. M. Menze, C. Heipke, and A. Geiger, "Object scene flow," ISPRS J. Photogram. Remote Sens. (JPRS), vol. 140, pp. 60–76, 2018.
4. C. Jiang, D. P. Paudel, D. Fofi, Y. Fougerolle, and C. Demonceaux, "Moving object detection by 3d flow field analysis," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 4, pp. 1950–1963, 2021.
5. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 2758–2766.
6. E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2462– 2470.
7. J. Y. Jason, A. W. Harley, and K. G. Derpanis, "Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness," in European Conference on Computer Vision. Springer, 2016, pp. 3–10.
8. Z. Ren, J. Yan, B. Ni, B. Liu, X. Yang, and H. Zha, "Unsupervised deep learning for optical flow estimation," in Proceedings of the AAAI Conference on Artificial Intelligence, 2017.
9. Y. Wang, Y. Yang, Z. Yang, L. Zhao, P. Wang, and W. Xu, "Occlusion aware unsupervised learning of optical flow," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 4884–4893.

10. S. Meister, J. Hur, and S. Roth, “Unflow: Unsupervised learning of optical flow with a bidirectional census loss,” in Proceedings of the AAAI Conference on Artificial Intelligence, 2018.
11. S. Alletto, D. Abati, S. Calderara, R. Cucchiara, and L. Rigazio, “Self-supervised optical flow estimation by projective bootstrap,” IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 9, pp. 3294–3302, 2018.
12. J. Janai, F. Guney, A. Ranjan, M. Black, and A. Geiger, “Unsupervised learning of multi-frame optical flow with occlusions,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 690–706.
13. P. Liu, I. King, M. R. Lyu, and J. Xu, “Ddflow: Learning optical flow with unlabeled data distillation,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 01, 2019, pp. 8770–8777.
14. P. Liu, M. Lyu, I. King, and J. Xu, “Selflow: Self-supervised learning of optical flow,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 4571–4580.
15. Y. Zou, Z. Luo, and J.-B. Huang, “Df-net: Unsupervised joint learning of depth and flow using cross-task consistency,” in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 36–53.
16. Ranjan, V. Jampani, L. Balles, K. Kim, D. Sun, J. Wulff, and M. J. Black, “Competitive collaboration: Joint unsupervised learning of depth, camera motion, optical flow and motion segmentation,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 12 240–12 249.
17. G. Wang, C. Zhang, H. Wang, J. Wang, Y. Wang, and X. Wang, “Unsupervised learning of depth, optical flow and pose with occlusion from 3d geometry,” IEEE Transactions on Intelligent Transportation Systems, 2020.
18. R. Jonschkowski, A. Stone, J. T. Barron, A. Gordon, K. Konolige, and A. Angelova, “What matters in unsupervised optical flow,” in Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer, 2020, pp. 557–572.
19. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2012, pp. 3354–3361.
20. M. Menze, C. Heipke, and A. Geiger, “Joint 3d estimation of vehicles and scene flow.” in ISPRS Workshop on Image Sequence Analysis (ISA), vol. 2, 2015.
21. J. J. Gibson, “The perception of the visual world.” Houghton Mifflin, 1950.
22. B. Horn and K. Berthold, “Schunck. determining optical flow,” Artificial Intelligence, vol. 17, no. 1-3, pp. 185–203, 1981.
23. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, “High accuracy optical flow estimation based on a theory for warping,” in Proc. Eur. Conf. Comput. Vis., 2004, pp. 25–36.
24. D. Sun, S. Roth, and M. J. Black, “Secrets of optical flow estimation and their principles,” in 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, 2010, pp. 2432–2439.
25. Ranjan and M. J. Black, “Optical flow estimation using a spatial pyramid network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 4161–4170.

26. G. Yang and D. Ramanan, “Volumetric correspondence networks for optical flow,” Proc. Adv. Neural Inf. Process. Syst., vol. 5, p. 12, 2019.
27. D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, “A naturalistic open source movie for optical flow evaluation,” in European conference on computer vision. Springer, 2012, pp. 611–625.
28. S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski, “A database and evaluation methodology for optical flow,” International journal of computer vision, vol. 92, no. 1, pp. 1–31, 2011.
29. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE transactions on image processing, vol. 13, no. 4, pp. 600–612, 2004.
30. Y. Wang, P. Wang, Z. Yang, C. Luo, Y. Yang, and W. Xu, “Unos: Unified unsupervised optical-flow and stereo-depth estimation by watching videos,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 8071–8081.
31. Y. Zhong, P. Ji, J. Wang, Y. Dai, and H. Li, “Unsupervised deep epipolar flow for stationary or dynamic scenes,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 12 095–12 104.
32. P. Liu, I. King, M. R. Lyu, and J. Xu, “Flow2stereo: Effective selfsupervised learning of optical flow and stereo matching,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 6648–6657.
33. G. Wang, H. Wang, Y. Liu, and W. Chen, “Unsupervised learning of monocular depth and ego-motion using multiple masks,” in 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019, pp. 4724–4730.
34. D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, “Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 8934–8943.
35. D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.