Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 9, Issue. 6, June 2020, pg.10 – 14

Evaluation Datasets and Benchmarks for Optical Flow Algorithms: A Review

Ritik Mathur[1]
[1]B.Tech, Electrical Engineering, Indian Institute of Technology Roorkee, Uttarakhand, India
[1] rmathur@me.iitr.ac.in

Abstract— Optical flow analysis for motion estimation, image segmentation, object detection and tracking has significantly revolutionized the fields of computer vision and robotics. Optical flow datasets provide a base and benchmark for training, testing and comparison of optical flow algorithms. This paper provides a review and analysis of available datasets that can be used for training and evaluation of optical flow algorithms in various applications of computer vision and robotics. In this paper, optical flow datasets are discussed with different attributes that can be utilized for comparison, reflecting the advantages and correct usage for key implementation of a specific task. In addition to this, open research challenges for the generation of finer and preferable optical flow datasets have been discussed.
Keywords— Optical flow, Datasets, Benchmark, Middlebury, MPI-Sintel, Flying Chairs, ChairsSDHom, KITTI, CrowdFlow, CreativeFlow+, FlyingThing3D, Monkaa, Driving, Computer vision, Robotics

I. INTRODUCTION

The motion of an object between successive frames of sequence, caused by the relative movement between the object and camera is termed as optical flow. The principle of optical flow is an involved research domain in machine vision systems and robotics for motion estimation, image segmentation, object detection and video compression. Recent extensions and modifications in optical flow analysis have considerably strengthened the domain of computer vision and robotics by supporting visual odometry[1] and robot navigation[2]. Optical flow[3] represented as sparse velocity vectors with point tracking as well as dense flow vectors per pixel over the entire frame enables estimation of rate of apparent motion between the viewer and object onto the two-dimensional image plane. The selection of an adequate and promising dataset is indispensable to train, compare, test and provide a benchmark for the proposed optical flow algorithm with the existing ones. Gathering annotated data for training and testing of optical flow algorithms is often time-consuming, expensive and is highly influenced by how an individual perceives the relative motion between an object and camera between consecutive frames of sequences. Size, complexity and diversity are major factors affecting the selection of a particular dataset as an evaluation measure for optical flow. The usability of an optical flow dataset for real-time applications is often highly influenced by the limitation of intricacy involved in the respective field.

Middlebury[4], MPI-Sintel[5] and KITTI 2012[6]/2015[7] are the four established optical flow datasets most commonly employed for evaluation. However, recent work in computer vision and robotics has opened a room for improvement and addition of better datasets to focus on a subset of issues and ensure better generalization. This necessitates more promising synthetic data for optical flow analysis, especially in high-level applications with minute details. Synthetic data often provides a room for benchmarking by employing additional demanding datasets with meticulous ground-truth information. In this paper, available datasets for optical flow analysis are explored to appreciate the major factors affecting the selection of a particular dataset as an evaluation measure for optical flow analysis.

II. RELATED WORK

1) Middlebury[4]: The Middlebury dataset and evaluation methodology, one of the most established and commonly used, contains 8 training and 8 test sequences with endpoint and angular error as measures of flow accuracy and interpolation and normalized interpolation error as measures of interpolation quality with an average and maximal velocities of about 4 and 22 respectively. Results can be submitted for evaluation in .flo format at publicly available benchmarking website. The dataset lacks the complexity of real-world with displacements typically below 10 pixels and therefore, it is not preferable for challenging real-time applications as well as for deep learning-based optical flow training and testing due to limited size, complexity and diversity.

2) CrowdFlow[8]: CrowdFlow dataset is recommended for mass-analysis applications such as estimation of movement of pedestrians, especially in highly crowded scenes. To simulate motion for visual crowd analysis, the Unreal Engine has been employed. The dataset consists of 10 sequences with lengths ranging between 300 and 450 frames with a frame rate of 25HZ and an HD resolution organized in continuous sequences allowing the evaluation of temporal consistencies unlike Middlebury[4] and KITTI 2012[6]/2015[7]. Ground truth flow data is provided as optical flow(foreground and background) and trajectories(dense and sparse) for each individual. In addition to this, a more discriminative trajectory-based long-term evaluation metric has been introduced to capture drifting and other estimation errors which are time-dependent. Crowd-tracking accuracy is compared with state-of-the-art optical flow methods on the proposed synthetic dataset and UCF crowd tracking dataset[9].

3) MPI-Sintel[5]: Based on an open-source 3D animated short movie ‘Sintel’, this richly varied and illuminated dataset has been introduced with an evaluation website for comparison and error statistics. Extensive comparison to Middlebury[4] dataset has been provided on basis of the difficulty of evaluation, sequence length, amount of data, image resolution, large motions, blur effect, motion boundaries and occluded regions, real-world challenges, transparency and ranking allowing one to select finer evaluation measure for optical flow training and testing. The dataset consists of 1064 training and 564 test frames in an 8-bit PNG format with a frame rate of 24 frames per second and an average and maximal velocities of 5 and 445 respectively acquired from 35 selected clips from the movie Sintel. The main disadvantage of the MPI-Sintel dataset is that these naturalistic video sequences are less realistic for real-time computations such as self-driving vehicle applications.

4) Flying Chairs[10] *and ChairsSDHom*[11]: Large scale datasets with ground truth information are crucial for supervised training in case of deep CNNs (Convolutional Neural Networks) which demand data augmentation and synthetic data. ‘Flying Chairs’, an artificial dataset with optical flow ground truth in the form of 22872 image pairs, has been proposed to train convolutional networks presented in FlowNet[10] with competitive accuracy by applying affine transformation parameters for the background and the chairs, which can be interpreted as planar relative motion between 3D chair models and random backgrounds from Flickr. To avoid overfitting and improve generalization, data augmentations such as geometric transformations as well as additive Gaussian noise have been implemented. “ChairsSDHom”, an artificial dataset to provide flow magnitude histograms close to those of the UCF101 dataset[12] with optical flow ground truth, has been proposed for training convolutional networks in FlowNet 2.0[11]. The dataset finds application on real-world data with small displacements as it is abstract enough to not overfit to any realistic scenario. The paper further discusses the idea that the order in which data is presented while training CNNs significantly affects the performance. These datasets do not handle training with non-rigid movements which is principal as non-simulated real objects do not always involve rigid motion.

5) KITTI 2012[6]/2015[7]: The KITTI 2012 and 2015 datasets and online benchmarks for stereo, optical flow, visual odometry[1], SLAM[13], 3D object detection and tracking involve diverse real-world traffic situations and is recommended for optical flow analysis in autonomous driving applications such as mobile and field robotics. The dataset is available in the form of 200 training scenes and 200 test scenes with 4 color images per scene saved in lossless PNG format. In addition, the development kit furnishes details about the data format and MATLAB/C++ utility functions for reading and writing disparity maps and flow fields. The movable recording platform used to capture this calibrated, synchronized and rectified autonomous driving dataset involves four video cameras (two color and two grayscale cameras), a rotating 3D laser scanner and a combined GPS/IMU inertial navigation system with an average and maximal velocities of 9 and 549 for KITTI 2012 and 8 and 724 pixels for KITTI 2015. The raw data set is divided into the categories 'Road', 'City', 'Residential', 'Campus' and 'Person' for diverse applications.

6) FlyingThings3D, Monkaa and Driving[14]: CNN-based (Convolutional Neural Networks) methods provide a promising approach for optical flow estimation by extracting deep, abstract and multi-scale features from input images at a much faster pace as compared to variational methods. For reasonable accuracy and precision, ground-truth optical flow annotated data is vital in sufficient number. A synthetic dataset containing 35000 stereo image pairs with ground truth disparity, optical flow, and scene flow has been proposed to generate a sufficiently large dataset for the training of convolutional neural networks for challenging vision tasks such as stereo, flow and scene flow estimation. The entire dataset suite is divided into three subsets with tens of thousands of frames animated using open source 3D creation suite Blender. Data augmentation techniques such as spatial and chromatic transformations have been utilized to introduce better variability in the training data . The applicability of the proposed synthetic dataset to scene flow estimation has been proved by training an end-to-end convolutional neural network, with disparity estimation 1000 times faster.
FlyingThings3D: This includes about 25000 stereo frames, fast and fully automatic generated data with dense accurate ground truth data of commonly used objects moving along randomized 3D paths. The base of each scene is derived from 200 static background objects with randomly scaled, rotated and textured cuboids and deformed cylinders.
Monkaa: Similar to the MPI Sintel dataset, Monkaa has been derived from an animated short film 'Monkaa' and is characterised with non-rigid motion for flow estimation. Random increments and changes to camera's position have been employed to generate additional new softly-articulated modifications to the scenes from the movie.
Driving: This dataset resembles the KITTI[7] dataset and is more preferable for optical flow analysis in dynamic driving applications. The sequences have been derived from car models similar to the FlyingThings3D dataset along with additional simple street lights and thoroughly particularized tree models from 3D Warehouse.

7) Creative Flow+ Dataset[15]: This multi-style artistic video dataset available in the form of 124K+ train set frames and 10K test set frames acts as an optical flow evaluation metric on non-photorealistic images and messy stylized content. 40 textured line styles have been randomly employed to create 3000 animated sequences for the dataset in addition to 38 shading styles implemented using Blender 2.79 python API and Stylit 3D contributing stylization algorithm. The dataset provides a high resolution of 1500x1500 and more diverse visual styles as compared to most available optical flow datasets and opens room for extensive research opportunities for optical flow estimation of non-photorealistic and stylized content. To combat size and memory issues, data compression to 570 GB with split downloading options are made available.

III. COMPARISON

Various datasets have been presented for the optical flow training and evaluation. A well organized and methodical comparison allows the selection of a preferable dataset for key implementation based on the complexity of a specific application. The comparison between different optical flow datasets and their key aspects have been listed in Table-1.

TABLE I
VARIOUS AVAILABLE OPTICAL FLOW DATASETS WITH THEIR KEY ASPECTS

Dataset	#Frames	Resolution	Year	Relevance
Middlebury	16	316 × 252 - 640 × 480	2007	Basic less demanding evaluation metric with limited complexity
MPI-Sintel	1628	1024 × 436	2012	Naturalistic richly-varied illuminated motion
KITTI 2012	778	1242 × 375	2012	Real-world autonomous driving applications
KITTI 2015	800	1242 × 375	2015	Real-world autonomous driving applications
CrowdFlow	3200	1280 × 720	2018	Visual crowd flow applications
Flying Chairs	22,872	960×540	2015	Train CNNs
Flying Things3D	4,248 (test) 21,818 (train)	960×540	2015	Train CNNs for visual applications

Monkaa	8,591	960×540	2015	Non-rigid softly articulated motion for CNNs
Driving	4392	960×540	2015	Dynamic viewpoint of a driving car for CNNs
Creative Flow+	10,031 (test) 124,390 (train)	1500×1500	2019	Non-photorealistic and stylized content-based applications

IV. OPEN RESEARCH CHALLENGES

1) Scalability: The ability of an optical flow dataset to be used or produced in a range of capabilities to adapt to increased demand for an application is termed as scalability of the dataset. Datasets with highly varying magnitudes, units and range offer complications particularly for mapping from input variables to an output variable in a deep learning neural network model for optical flow calculation[16]. Data preparation techniques such as data normalization and standardization are often employed for feature scaling to improve stability and performance.

2) Time span: Recent developments in computer vision, robotics and image processing have necessitated the availability of an optical flow dataset with more diversity and complexity to create a room for differentiation of proposed algorithm with existing ones in a better and robust manner. However, each dataset is preferable to manipulate for a limited duration governed by the obtainable resources in hand and therefore, time span for dataset matters for flawless implementation and evaluation to remain competitive in an increasingly digital world.

3) Memory and Computational cost: Artificial intelligence and deep learning systems for optical flow analysis rely on massive data exceedingly. The amount of usable data that a memory unit can access and process in a given period affects inactivity on to the data bus. Boosting memory performance to handle a rising flood of data is an active research domain in the semiconductor design industry.

4) Diversity and Complexity: One of the major aspects of a diverse optical flow dataset is application-specific complexity and a trade-off to avoid high bias and overfitting of an algorithm. This ensures satisfactory generalization of the trained approach to tackle real-world analysis with acceptable results on unseen data.

5) Ground Truth Reliability: The ground truth flow data affects the predictive capability of an optical flow algorithm by impacting consistency, completeness and accuracy. Reliability and trustworthiness of available ground truth flow information are crucial for the establishment of a robust base for training and evaluation of optical flow algorithms for motion estimation, object detection and object tracking tasks in machine vision and robotics.

V. CONCLUSION

This paper presents a review on existing datasets available for training and evaluation of optical flow algorithms. Motion estimation with optical flow has a great potential for research and development in computer vision, image processing and robotics. Several key aspects which are considered in the review can help in better understanding the potential and correct selection of dataset for key implementation for specific applications to ensure cost and computational effectiveness. Future works and researches will facilitate in adding finer and preferable base as well as best practices for comparison and evaluation of optical flow algorithms to avoid saturation and limitations.

REFERENCES

[1] K. Yousif, A. Bab-Hadiashar, and R. Hoseinnezhad, "An Overview to Visual Odometry and Visual SLAM: Applications to Mobile Robotics," *Intelligent Industrial Systems*, 2015, doi: 10.1007/s40903-015-0032-7.

[2] H. Chao, Y. Gu, and M. Napolitano, "A survey of optical flow techniques for robotics navigation applications," *Journal of Intelligent and Robotic Systems: Theory and Applications*, 2014, doi: 10.1007/s10846-013-9923-6.

[3] F. Becker, S. Petra, and C. Schnörr, "Optical flow," in *Handbook of Mathematical Methods in Imaging: Volume 1, Second Edition*, 2015.

[4] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, 2011, doi: 10.1007/s11263-010-0390-2.

[5] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012, doi: 10.1007/978-3-642-33783-3_44.

[6] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, doi: 10.1109/CVPR.2012.6248074.

[7] M. Menze and A. Geiger, "Object scene flow for autonomous vehicles," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, doi: 10.1109/CVPR.2015.7298925.

[8] G. Schroder, T. Senst, E. Bochinski, and T. Sikora, "Optical Flow Dataset and Benchmark for Visual Crowd Analysis," in *Proceedings of AVSS 2018 - 2018 15th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2019, doi: 10.1109/AVSS.2018.8639113.

[9] S. Ali and M. Shah, "Floor fields for tracking in high density crowd scenes," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2008, doi: 10.1007/978-3-540-88688-4-1.

[10] A. Dosovitskiy *et al.*, "FlowNet: Learning optical flow with convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, doi: 10.1109/ICCV.2015.316.

[11] 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 *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, doi: 10.1109/CVPR.2017.179.

[12] A. R. Z. and M. S. Khurram Soomro, "UCF101 - Action Recognition Data Set," *www*, 2012. .

[13] C. Stachniss, J. J. Leonard, and S. Thrun, "Simultaneous localization and mapping," in *Springer Handbook of Robotics*, 2016.

[14] N. Mayer *et al.*, "A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, doi: 10.1109/CVPR.2016.438.

[15] M. Shugrina et al., "Creative flow+ dataset," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019, doi: 10.1109/CVPR.2019.00553.

[16] Z. Ren, J. Yan, B. Ni, B. Liu, X. Yang, and H. Zha, "Unsupervised deep learning for optical flow estimation," in *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 2017.

