Bridging the Gap: Optical Flow Models in Real-World Scenarios

Marijn Timmerije
m.timmerije@student.tudelft.nl

How do existing optical flow models perform in real-world scenarios with large displacements of 20 pixels or more?

Supervisors: Jan van Gemert
Sander Gielisse

Background

Optical Flow Estimation:

- Critical for many computer vision tasks like autonomous driving an robotics.

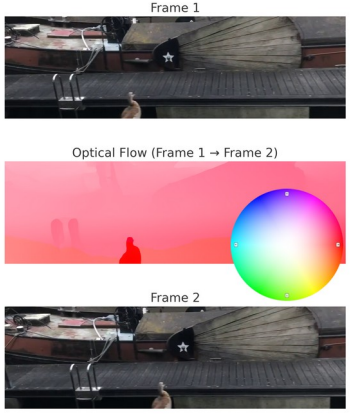
Recent Developments:

- Integration of deep learning resulted in large improvements.

Knowledge Gap:

- Models are generally benchmarked on synthetic datasets and datasets that indiscriminately combine confounders such as non-rigid motion, light changes, etc.
- Dataset needs to be generated and models benchmarked to verify improvements in space.

Fig 1: Example of Optical Flow Visualization



Methodology

Tool Development:

- Developed a tool to annotate frames to create dataset, in existing (KITTI [2]) format.

Data Generation:

- Record suitable footage with large displacements.
- Sort into different categories (Fig 3 and 4)
- Annotate ground truth using tool (Fig 2)

Select Metrics:

- EPE, Fl-all and Accuracy
- Fl-all over distances and Distance-binned EPE

Model Selection:

- RAFT[1] as baseline, as it was SOTA in 2020.
- Models that have reported good performance on large displacements on other datasets.
  - FlowFormer++, GMFlow, GMFlowNet, MemFlow, MemFlow-T

Evaluation:

- Evaluate pretrained models on relevant statistics
- Compare performance of models (Fig 5, 6)

Fig 2: Screenshot of Annotation Tool

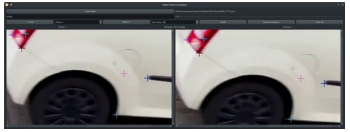


Fig 3: Datasets Composition

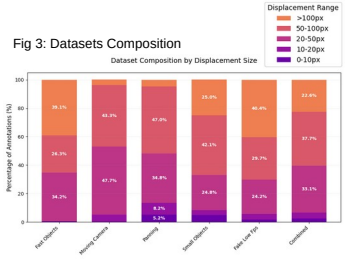
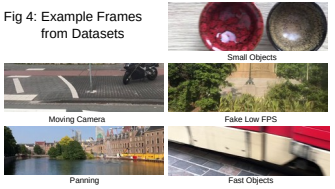


Fig 4: Example Frames from Datasets



Results and Discussion

Best Performance: MemFlow-T [3] outperformed competition by almost every metric. FlowFormer++ also performed quite well consistently (Fig 2 and 3).

Unexpected Results: On specific datasets, MemFlow performed better than MemFlow-T, so its vision transformer does not always improve results.

Data Sparsity: Because small (<10px) and extreme (>100px) displ. were not prevalent in (every) dataset, the statistics surrounding it have low confidence (Fig 3, 6)

Transformer-based architectures: These seem to result in large improvements for large displacements.

Future Work

Larger Dataset: Dataset is only n=250, because of technical issues. Further research needs larger datasets for better statistical confidence.

Dynamic Scenes: Models should be ran against more dynamic scenes that still avoid other confounders.

Cross-data Generalization: Benchmarking should be done against other pretraining too.

Fig 5: Outlier and Accuracy Percentages

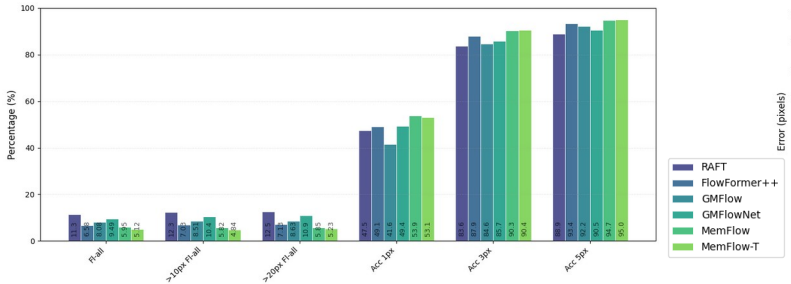
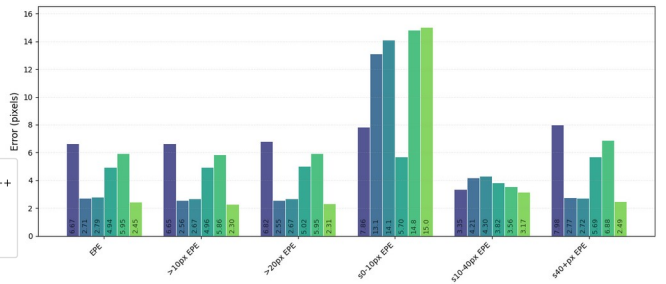


Fig 6: Endpoint Errors



References

[1] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In Proceedings of the European Conference on Computer Vision (ECCV), 2020.

[2] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In Conference on Computer Vision and Pattern Recognition (CVPR), 2015

[3] Qiaole Dong and Yanwei Fu. Memflow: Optical flow estimation and prediction with memory. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,