Comparing Optical Flow models on repetitive patterns in real-world images

Author
Jesse Klijnsma
J.B.Klijnsma@student.tudelft.nl

Responsible professor
Jan van Gemert

Supervisor
Sander Gielisse

1 BACKGROUND

- Optical flow estimation aims to find the 2D **pixel-wise motion** between frames.
- State-of-the art deep neural network models are generally trained on **synthetic data**.
- When models are trained and evaluated on real-world data, the ground-truth is based on other approximation algorithms.
- One of the challenges faced by the models is repetitive patterns, where different regions look visually identical.
- Models like **LiteFlowNet3** and **Ef-RAFT** claim that they have taken measures specifically against repetitive patterns

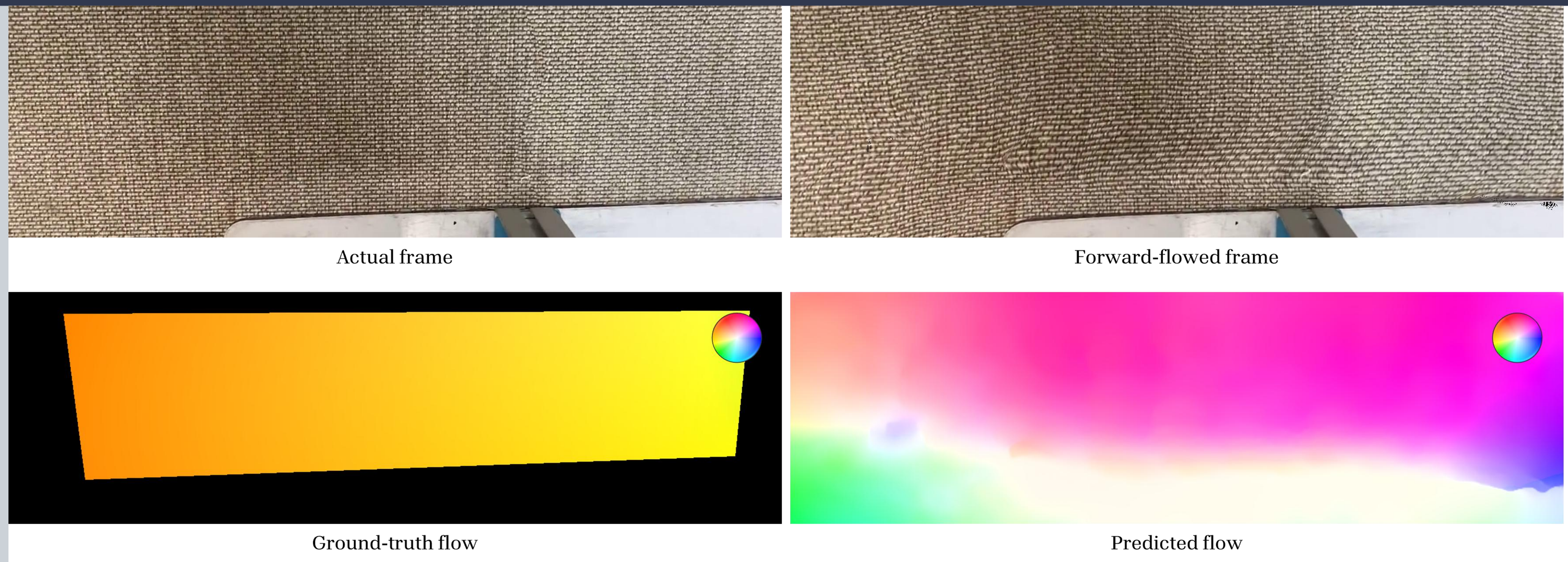

Figure 1: Example of a faulty prediction where the reconstructed second frame looks visually similar to the actual second frame

4 RESULTS

Checkpoint	Things		Sintel		Kitti		Mix	
	EPE	F1-All	EPE	F1-All	EPE	F1-All	EPE	F1-All
Model								
ccmr_p	-	-	0.4007	0.00%	0.3914	0.00%	-	-
ms_raft_p	-	-	-	-	-	-	0.4000	0.07%
dpflow	0.4069	0.07%	0.4021	0.07%	0.3949	0.00%	-	-
splatflow	-	-	-	-	0.4234	0.07%	-	-
ccmr	-	-	0.4521	1.09%	0.4087	0.00%	-	-
rpknet	0.4380	0.07%	0.4315	0.07%	0.4363	0.00%	-	-
csflow	0.4300	0.22%	-	-	0.4672	0.22%	-	-
lcv_raft	0.4525	0.07%	-	-	-	-	-	-
gmflownet_mix	0.4624	0.22%	0.4472	0.30%	-	-	-	-
irr_pwcnet_irr	0.4819	0.07%	-	-	-	-	-	-
liteflownet3s	-	-	0.4850	0.01%	-	-	-	-
flowformer_pp	0.4203	0.00%	0.4474	0.22%	0.6734	1.20%	-	-
unimatch_sc2	0.5061	0.01%	0.5075	0.00%	-	-	0.5329	0.00%
gmflow_p_sc2	0.5061	0.01%	0.5075	0.00%	-	-	0.5329	0.00%
rapidflow	0.4817	0.15%	0.4693	0.37%	0.6068	0.80%	-	-
unimatch_sc2_ref6	0.6606	2.19%	0.4476	0.25%	0.6214	1.11%	0.4111	0.00%
gmflow_p_sc2_ref6	0.6606	2.19%	0.4476	0.25%	0.6214	1.11%	0.4111	0.00%
Middle values omitted to maintain readability								
flownetsd	8.1114	42.00%	-	-	-	-	-	-
maskflownet	-	-	3.4979	6.90%	18.1352	25.44%	-	-
scopeflow	0.4950	0.15%	0.4665	0.67%	43.3554	50.93%	-	-
vcn	27.8040	26.19%	8.7887	15.85%	29.4939	50.95%	-	-
vcn_small	33.9843	26.50%	-	-	-	-	-	-
hd3	0.5905	1.46%	1.5735	1.98%	129.1069	57.52%	-	-

Table 1: End-Point-Error and F1-All scores for all models and checkpoints

2 RESEARCH QUESTION

How does the performance of optical flow prediction models compare on repetitive patterns in real-world footage?

1. Which models are most resilient to repetitive patterns in terms of End-Point-Error?
2. Does a low reconstruction error and a high EPE indicate a failure due to repetitive patterns?
3. Which models perform best according to the False Correspondence Index?

3 METHOD AND TOOLS

- **Data:** A dataset containing pairs of 24 image pairs is collected with a focus on repetitive patterns, such as tiled floors, brick walls, fences and textiles.
- **Annotation:** Using a special annotation tool, optical flow vectors will be sparsely annotated and stored into the KITTI'15 [1] format for evaluation. When the motion of a certain pattern only consists of an affine transform, it is possible to interpolate the flow between annotated points using homography, making the final annotation semi-dense.
- **Evaluation:** Using PTLFlow, a collection of models, 69 models with in total 159 model-checkpoint combinations are put under evaluation. The resulting flow predictions will be evaluated using End-Point-Error and F1-All, only where a ground-truth annotation exists.

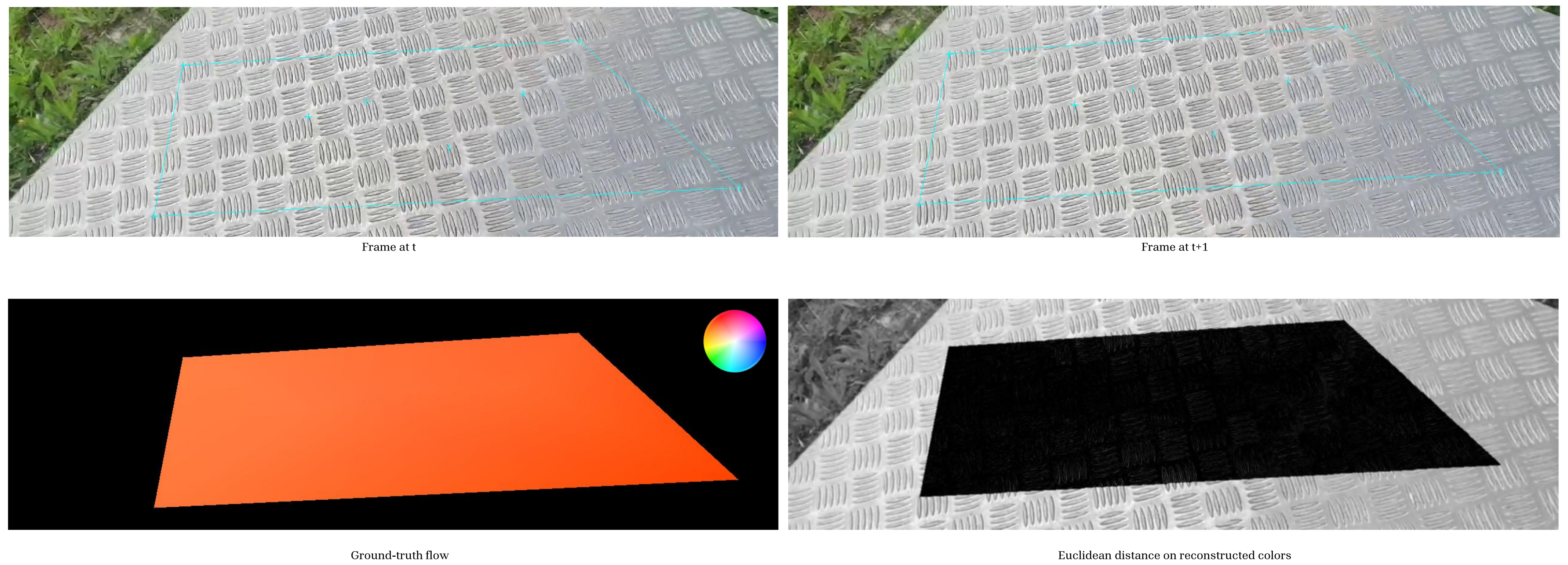

Figure 2: On the top row two subsequent frames and on the bottom row the ground-truth optical flow interpolated using homography and the error of this interpolation by reconstructing the second frame.

5 CONCLUSION

- **CCMR+** performs best overall on repetitive patterns in terms of End-Point-Error, although several other models perform very similarly
- A low reconstruction error and a high EPE indicate a failure due to repetitive patterns, as shown in figure 1.
- **CCMR+** performs best according to the False Correspondence Index, showing the strongest resilience against repetitive patterns. **MS-RAFT+**, **CCMR**, **DPFlow**, and **LiteFlowNet3s** follow closely, with very similar performance

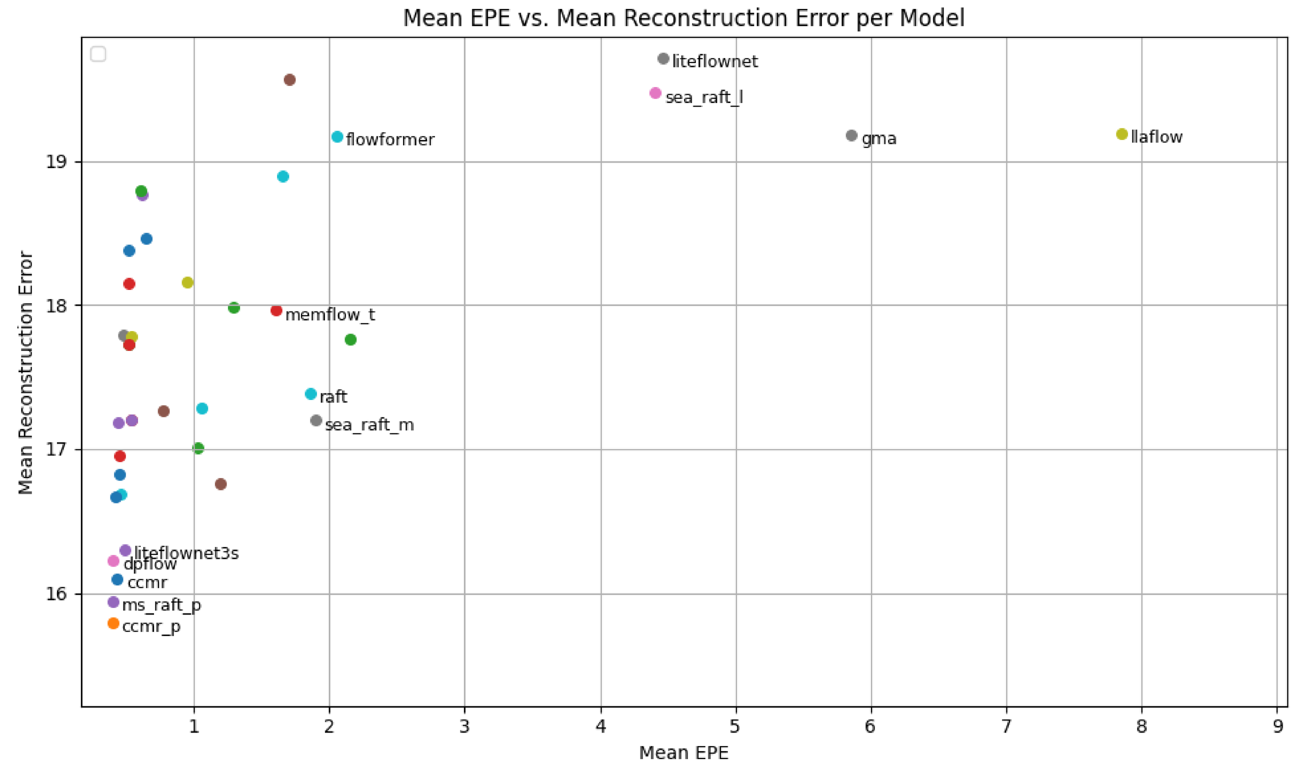

Figure 3: False Correspondence Index per model. Showing best performing models in the bottom left.

6 FUTURE WORK

- **Benchmark runtime performance:** The models are only evaluated on their accuracy, while runtime is also a large factor for selecting a model.
- **Larger dataset:** The current dataset is limited to 24 scenes. Including more dynamic and diverse scenes would lead to stronger evidence.

TUDelft