Learning Motion Patterns in Videos

Pavel Tokmakov Karteek Alahari Cordelia Schmid

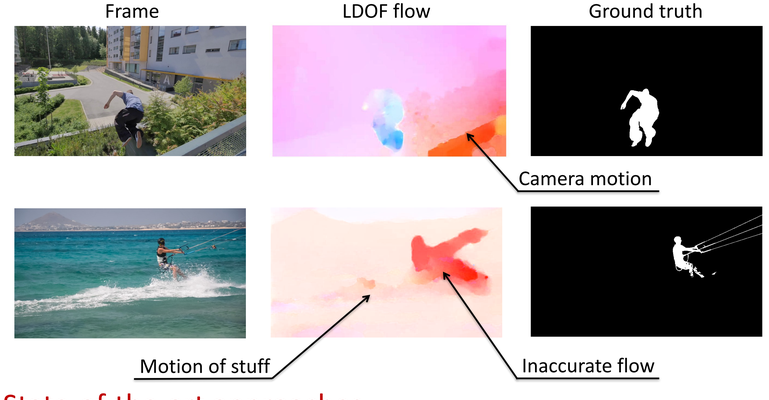
CVPR July 21-26 2017 HONOLULU

Code available: http://thoth.inrialpes.fr/research/mpnet

Motion segmentation

The problem of segmenting independently moving objects in videos.

Challenges



State-of-the-art approaches

Clustering dense point trajectories
[Brox and Malik, ECCV'10],
[Keuper et al., ICCV'15]

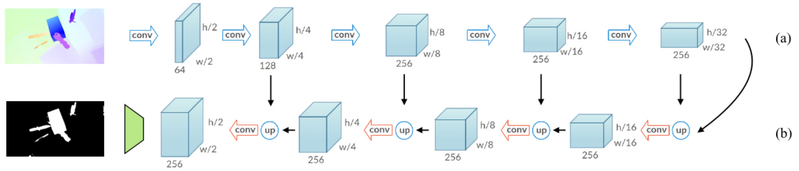
Heuristic optical flow-based methods
[Papazoglou and Ferrari, ICCV'13],
[Faktor and Irani, BMVC'14]

Neither of them uses learning to detect motion patterns

- We propose a model (MP-Net) for learning to segment independent motion
- Apply post-processing to handle the remaining challenges

Our approach: MP-Net

- CNN with an encoder (a) and a decoder (b) parts
- Takes optical flow as input and outputs an estimate of motion segmentation
- Learns to capture patterns that correspond to independent motion



Training data

- No dataset of real videos with dense pixel-level annotations is available
- We utilize synthetic data FT3D [Mayer et al., CVPR'16]
- FT3D provides ground truth flow, camera pose and instance segmentation
- We compute moving object labels from this data

Detecting motion patterns in real videos

- Post-processing to handle stuff in motion and flow inaccuracies
- Extract object proposals with SharpMask [Pinheiro et al., ECCV'16]
- Aggregate them into an objectness map to suppress motion of stuff
- Dense CRF [Krähenbühl et al., NIPS'11] for boundary refinement

Ablation studies

Experiments on FT3D, DAVIS and BMS-16

# dec.	Trained on FT3D with ...	FT3D	DAVIS
1	RGB single frame	68.1	12.7
	RGB pair	69.1	16.6
	GT flow	74.5	44.3
	GT angle field	73.1	46.6
	RGB + GT angle field	74.8	39.6
	LDOF angle field	63.2	38.1
4	GT angle field	85.9	52.4

Mean IoU on FT3D and DAVIS

- Flow is necessary for domain transfer
- Flow quality is important during training

Variant of our method	Flow used	Mean IoU
MP-Net	LDOF	52.4
MP-Net	EpicFlow	56.9
MP-Net + Objectness	LDOF	63.3
MP-Net + Objectness	EpicFlow	64.5
MP-Net + Objectness + CRF	LDOF	69.7
MP-Net + Objectness + CRF	EpicFlow	68.0

Results on DAVIS

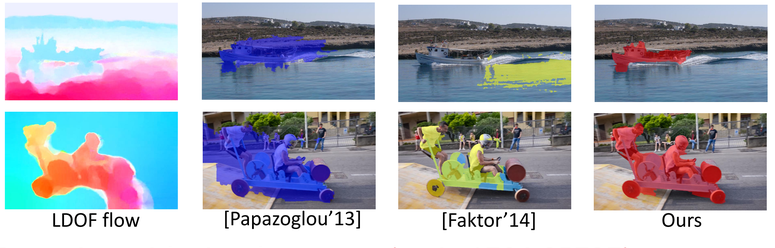
- Better flow in test helps MP-Net
- Post-processing is essential for top results
- It also cancels out the flow difference

Comparison to the state-of-the-art

Measure	[Faktor'14]	[Taylor'15]	[Brox'10]	[Lee'11]	[Papazoglou'13]	Ours
IoU	64.1	51.4	54.3	56.9	57.5	**69.7**
Boundary acc.	59.3	49.0	52.5	50.3	53.6	**66.3**
Temp. stab.	35.6	24.3	25.0	**19.0**	27.6	68.6

Results on DAVIS (see paper for full table)

Our frame-level method outperforms video-level approaches on DAVIS



Extension with visual memory (arxiv:1704.05737)

- An appearance stream to encode semantic information
- A visual memory module (ConvGRU) to segment objects after they stop