ECCV 2016 poster presentation

Motion Representation with Acceleration Images

Hirokatsu Kataoka[1], Yun He[1,2], Soma Shirakabe[1,2], Yutaka Satoh[1,2]

[1]Computer Vision Research Group, AIST, Japan [2]Human-Centered Vision Lab., University of Tsukuba, Japan

How to improve motion-based features?

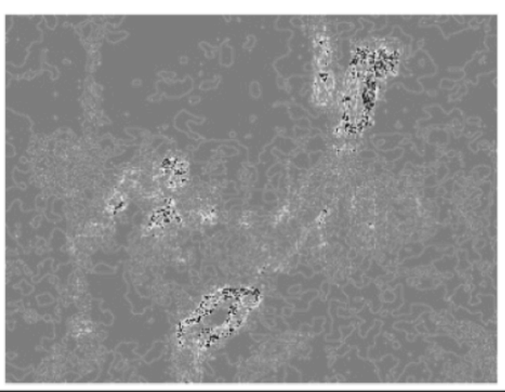
- Information of time differentiation is extremely important for a motion representation
  - The crosspoint of the IDT and the two-stream CNN is the strongest approach
  - 94.2% (TSN[1]) on the UCF101
  - It's important to enhance motion representation effectively

- Richer image representation other than position (RGB) and speed (flow) is needed

Acceleration-stream

- We employ acceleration-stream in addition to the spatial- and temporal-stream
  - The acceleration images are generated by differential calculations from a sequence of flow images
  - CNN can automatically catch an effective motion feature from sparse and noisy acceleration images



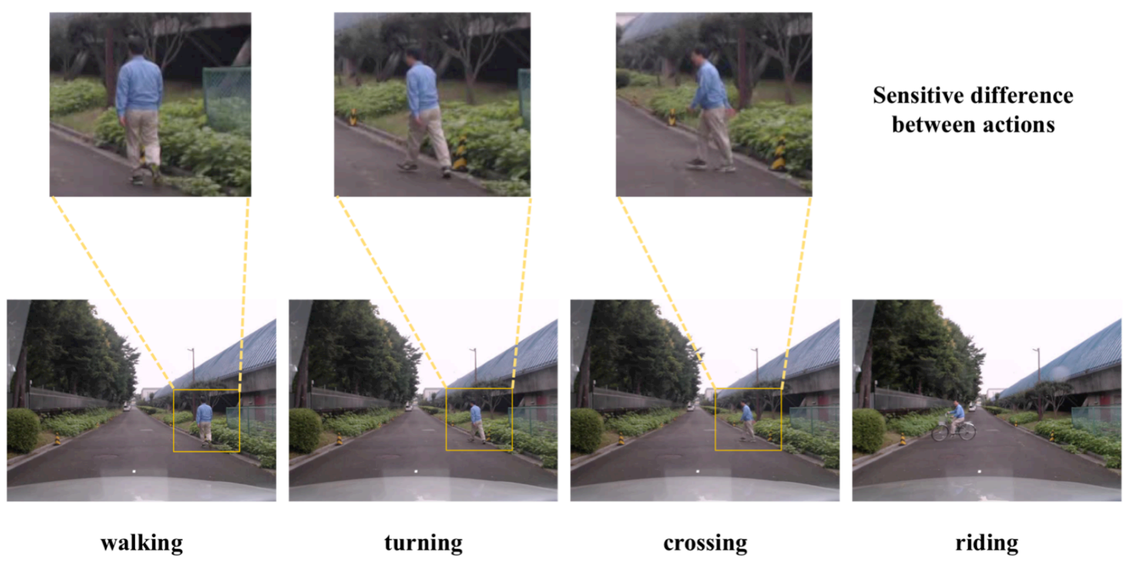


Physics quantity	I	I' (1st-order diff.)	I'' (2nd-order diff.)
Input	RGB	Flow image	Acceleration image
Stream	Spatial-stream	Temporal-stream	Acceleration-stream

Comparison

- Baseline: Very deep two-stream CNN [2]
  - 16-layer, UCF101 pre-trained model
  - Highly dropout ratio in FC layers of the acceleration stream

- Dataset: NTSEL [3]
  - 100 videos of pedestrian actions, *walking, turning, crossing,* and *riding a bicycle*
  - Each of the four actions has 25 videos: 15 for training and the other 10 for testing



Approach	% on NTSEL
Spatial stream	87.5
Temporal stream	77.5
Acceleration stream	82.5
Two-streams(S+T)	87.5
Three stream(S+T+A)	**90.0**

Conclusion

- We add acceleration stream to two-stream CNN
  - The representation of acceleration is different from RGB and flow
  - CNN can pick up necessary feature in the acceleration images

- The motion feature of acceleration is effective for action recog.
  - The result with stream(A) is better than stream(T)!

Reference

[1] L. Wang, et al. Temporal segment networks: Towards good practices for deep action recognition. in ECCV, 2016.
[2] L. Wang et al. Towards Good Practices for Very Deep Two-Stream ConvNets. In arXiv, 2015.
[3] H. kataoka, et al. walking activity recognition via driving recorder dataset. In ITSC,2015.