

Activity Normalization for Activity Detection in Surveillance Videos



Takashi Hosono¹, Kiyohito Sawada², Youngqing Sun¹,
Kazuya Hayase¹, Jun Shimamura¹

1 NTT Media Intelligence Laboratories, Japan

2 National Police Academy, Japan



Activity detection in surveillance videos

- Detects spatial position and time at which activity occurs

Example of ACTEV/VIRAT dataset



○ : activities

□ : objects relevant to activities

General processing flow

1. **Generates activity-proposal** on basis of object detection and tracking
2. **Estimates activity class** for each activity-proposal



General processing flow

1. **Generates activity-proposal** based on object detection and tracking
2. **Estimates activity class** for each activity-proposal



Difficulty of activity classification


- Diversity of activity appearances in the same class
 - caused by diversity of **object-movement directions** and **inter object positional relationships**

Diversity of object-movement direction

Class: carrying

Various movement directions



 : movement direction

Diversity of inter-object positional relationship

Class: loading

Various positional relationships



 : car  : person

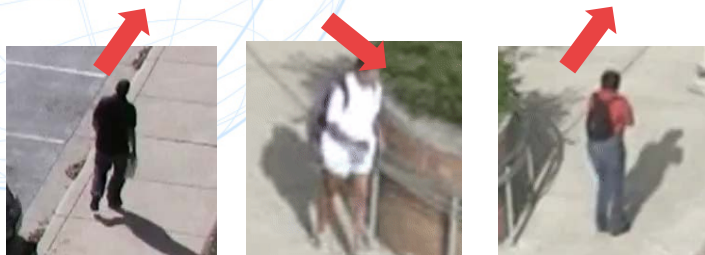
Approach: activity normalization

- **Align object-movement direction and inter-object positional relationships by rotating and flipping activity proposal**

Object-movement-direction normalization

Inter-object-positional-relationship normalization

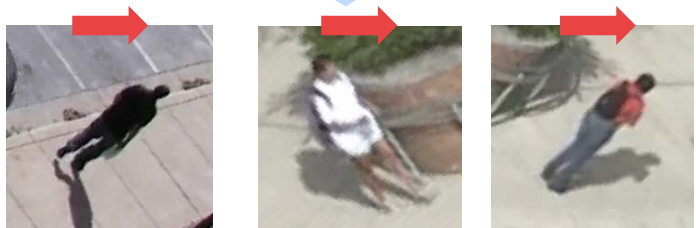
Before



Before



After



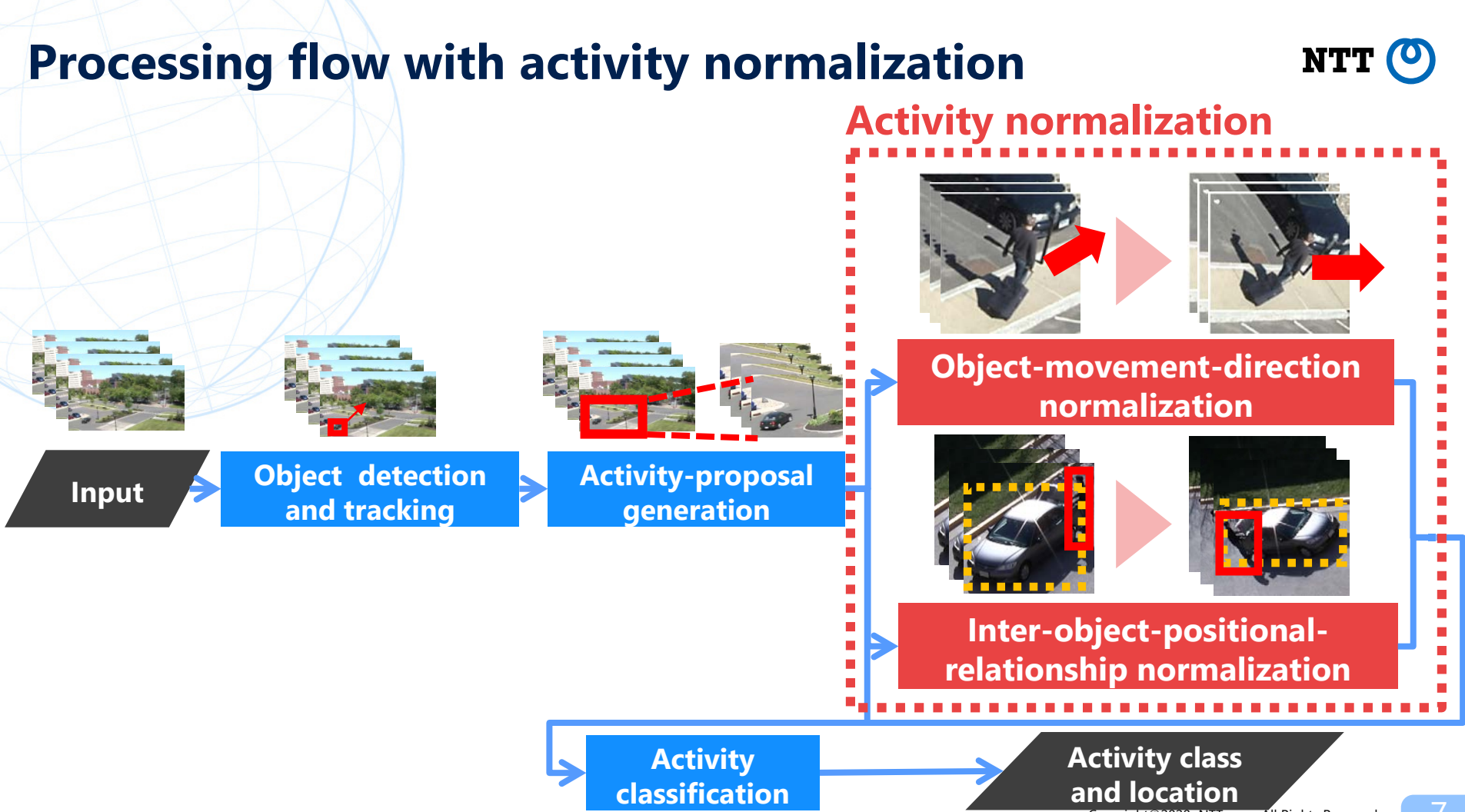
After



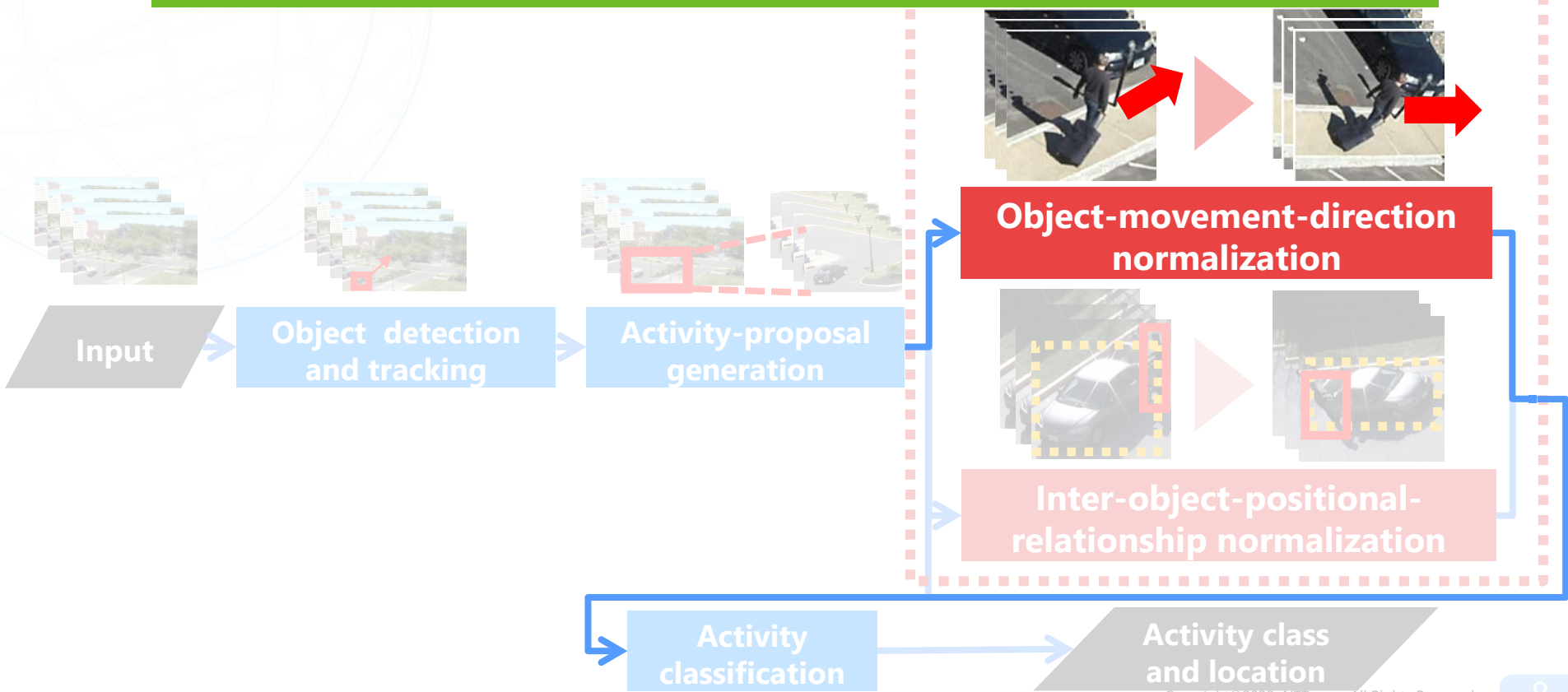
 : movement direction

 : car  : person

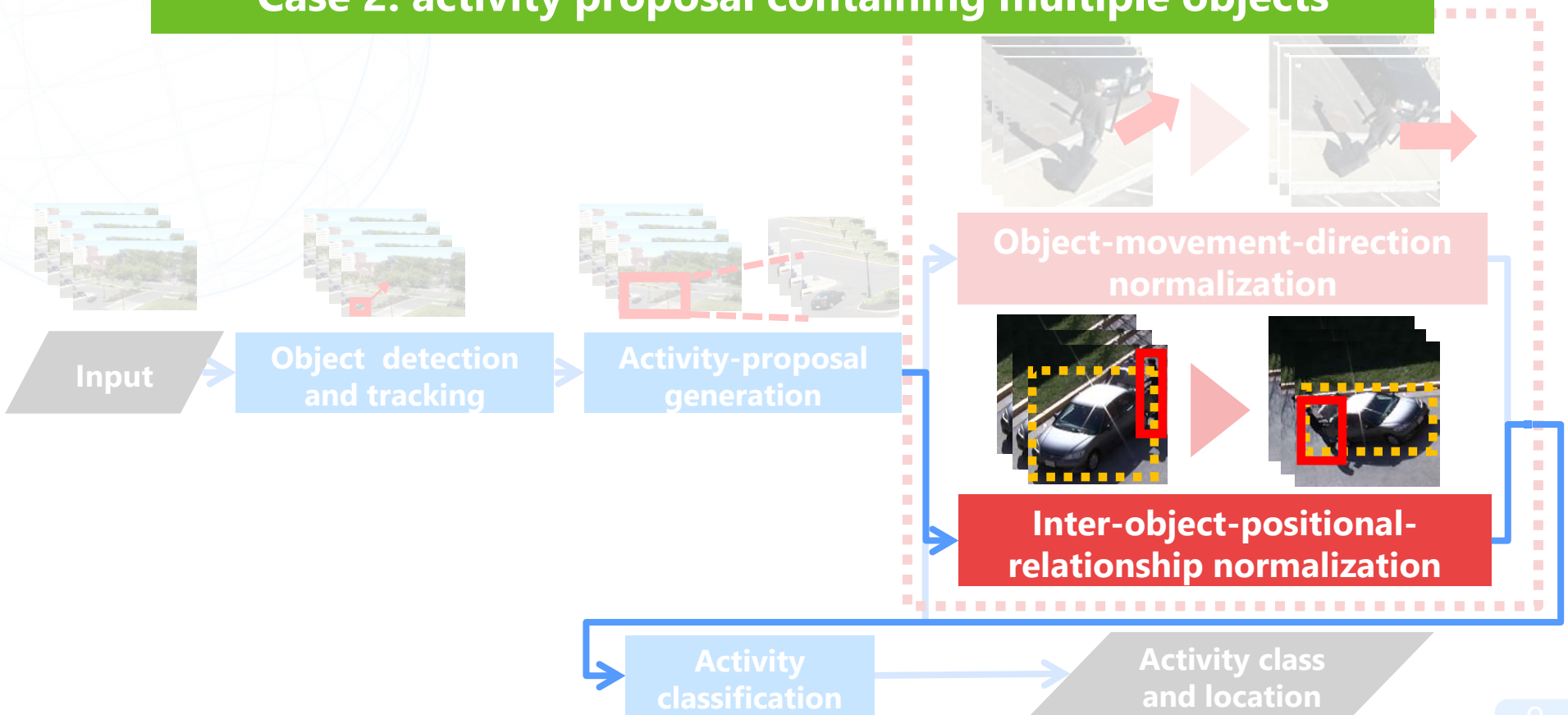
Processing flow with activity normalization



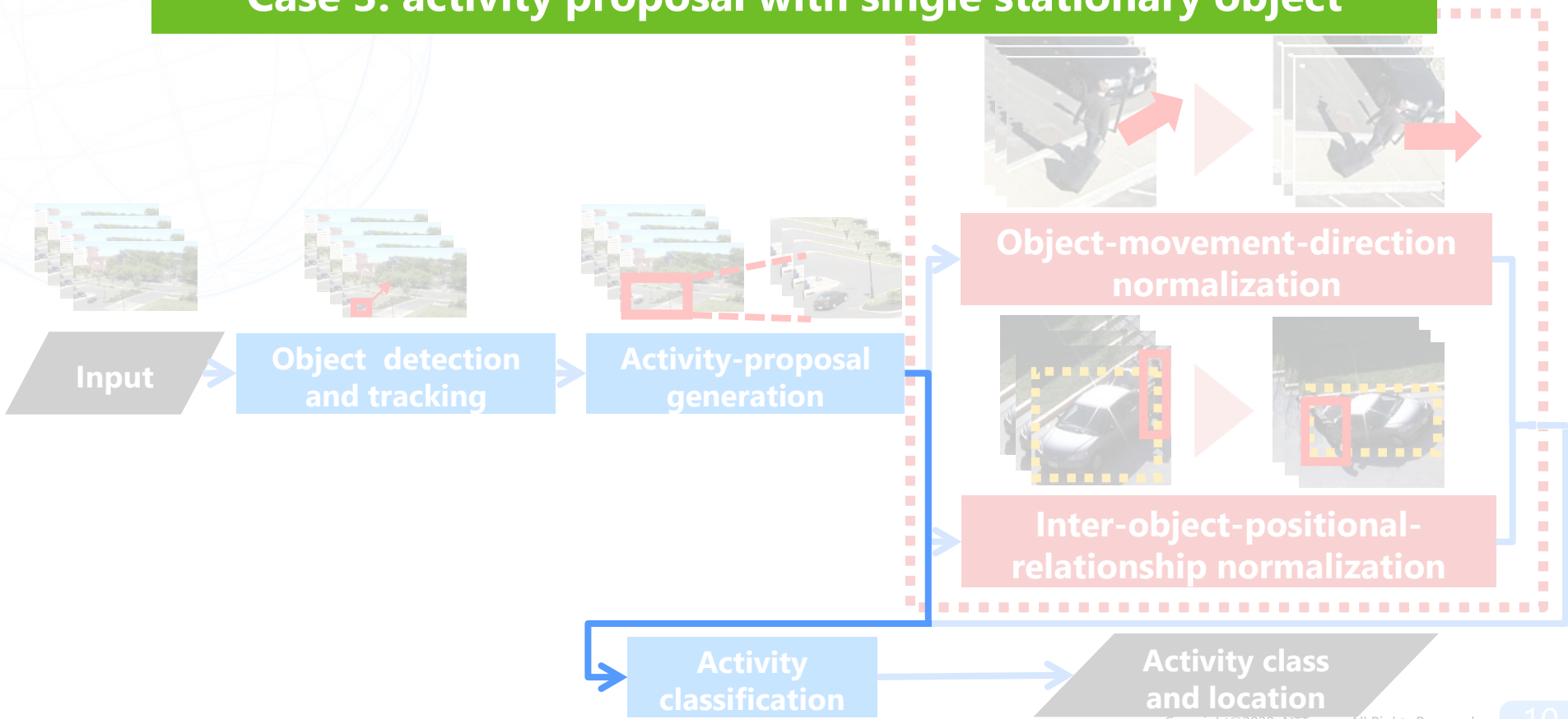
Case 1: activity proposal with a large amount of movement



Case 2: activity proposal containing multiple objects



Case 3: activity proposal with single stationary object




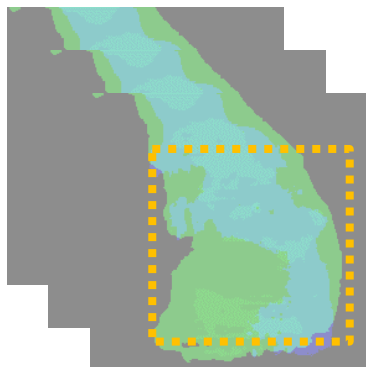
Object-movement-direction normalization

1. Calculate optical flow angle in object regions
2. Mode angle in object region is selected as object-movement direction
3. Activity proposal is rotated so that it becomes a fixed angle

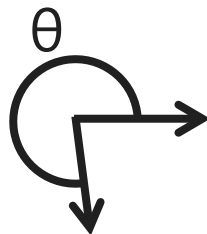


Activity proposal

: object



Optical flow



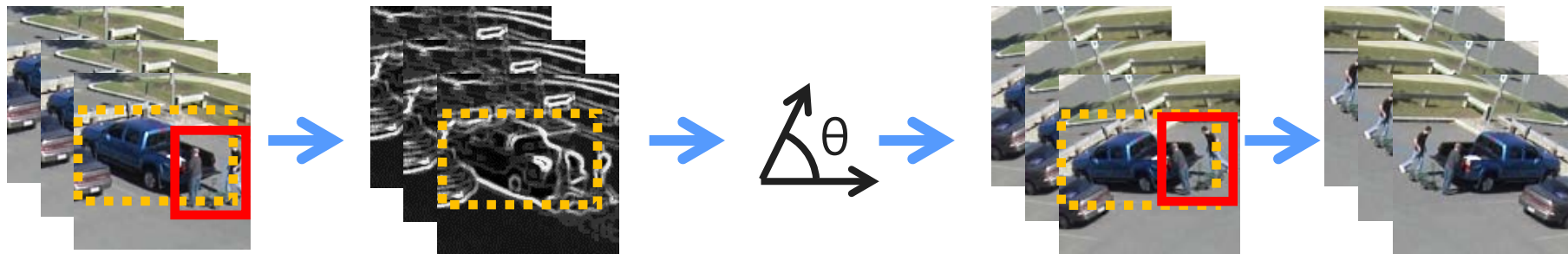
Object-movement
direction



Normalized activity
proposal

Inter-object-positional-relationship normalization

1. Calculate gradient vectors by Sobel filter
2. Mode angle in reference object region is selected as reference object direction
3. Activity proposal is rotated so that it becomes a fixed angle
4. Flip so that the left-right positional relationship is constant



Activity proposal

Gradient images

Reference object
direction

Normalized activity
proposal

 : car  : person

Evaluation

- Test with ActEV/VIRAT dataset
 - We evaluated only activity classification
 - ✓ Activity proposal is generated using ground truth
- Comparison methods
 - Baseline [Sun+ TRECVID19]
 - Baseline + data-augmentation (DA)
 - ✓ Activity proposal rotated and flipped in 16 directions
- **Mean precision improved by 0.05 with activity normalization**

Activity	Precision		
	Baseline	Baseline+DA	Ours
Vehicle turning right	0.682	0.622	0.827
Vehicle turning left	0.609	0.574	0.808
Vehicle U-turn	0.458	0.483	0.646
Activity carrying	0.950	0.904	0.933
Transport heavy carry	0.672	0.412	0.605
Pull	0.707	0.667	0.715
Riding	0.933	0.939	0.935
Loading	0.437	0.429	0.608
Unloading	0.251	0.381	0.279
Open trunk	0.243	0.292	0.147
Closing trunk	0.116	0.100	0.131
Opening	0.307	0.358	0.318
Closing	0.362	0.420	0.428
Entering	0.358	0.380	0.466
Exiting	0.384	0.519	0.468
Talking	0.774	0.784	0.798
Talking phone	0.043	0.035	0.028
Texting phone	0.003	0.005	0.004
Mean	0.461	0.461	0.508

Normalized activity proposals

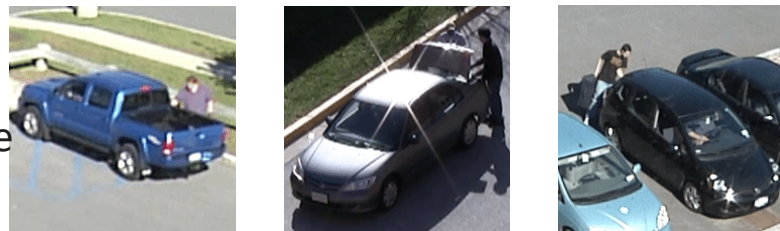
Class: vehicle turning right

Class: loading

Before



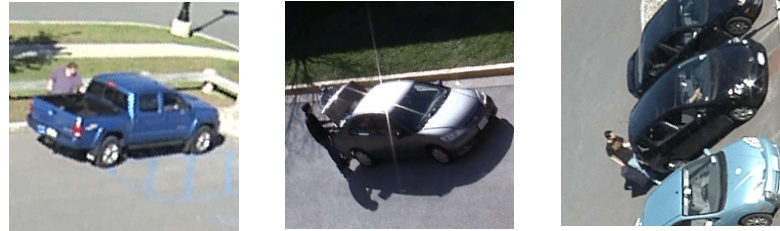
Before



After



After



Effects of two type normalization

Object-movement-direction normalization (baseline + mov)
is more effective for activity with movement

Method	mAP (car activity)	mAP (car-person activity)	mAP
Baseline	0.583	0.307	0.461
Baseline + mov	0.738	0.321	0.485
Baseline + pos	0.644	0.345	0.484
Baseline + mov + pos (ours)	0.760	0.356	0.508

Effects of two types normalization

Inter-object-positional-relationship normalization (baseline + pos) is more effective for activity with car-person interaction

Method	mAP (car activity)	mAP (car-person activity)	mAP
Baseline	0.583	0.307	0.461
Baseline + mov	0.738	0.321	0.485
Baseline + pos	0.644	0.345	0.484
Baseline + mov + pos (ours)	0.760	0.356	0.508

Test with whole pipeline

- Evaluated accuracy of activity detection with nAUDC (normalized partial area under the detection error trade-off curve)
 - Effect of our method is smaller than when ground truth activity proposals are used
- ➔ Problems remain when using realistic activity proposals

nAUDC (lower is better)

Method	nAUDC
Baseline [Sun+ TRECVID19]	0.589
Baseline + mov. + pos. (ours)	0.579

- Summary

- We proposed an **activity normalization** method to suppress the number of activity appearances
 - ✓ activity proposal is rotated and flipped so that the **object-movement direction** and **inter-object-positional relationship** are constant
- Experimental results showed that our method can improve activity classification accuracy

- Future work

- Make activity normalization method more robust to realistic activity proposals
- Validate its compatibility with other activity detection approaches



Thank you for watching