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

**Diversity of inter-object positional relationship**

Class: loading

Various positional relationships

- car
- person

: movement direction
**Approach: activity normalization**

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

**Object-movement-direction normalization**

Before

After

**Inter-object-positional-relationship normalization**

Before

After

: movement direction

: person

: car
Processing flow with activity normalization

Input

Object detection and tracking

Activity-proposal generation

Activity classification and location

Activity normalization

Object-movement-direction normalization

Inter-object-positional-relationship normalization
Processing flow with activity normalization

Case 1: activity proposal with a large amount of movement

Input

Object detection and tracking

Activity-proposal generation

Activity classification

Object-movement-direction normalization

Inter-object-positional-relationship normalization

Activity class and location
Case 2: activity proposal containing multiple objects

Processing flow with activity normalization

Input

Object detection and tracking

Activity-proposal generation

Object-movement-direction normalization

Inter-object-positional-relationship normalization

Activity classification

Activity class and location
Processing flow with activity normalization

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
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
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

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

Class: vehicle turning right

Before

After

Class: loading

Before

After
## Effects of two type normalization

<table>
<thead>
<tr>
<th>Method</th>
<th>mAP (car activity)</th>
<th>mAP (car-person activity)</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.583</td>
<td>0.307</td>
<td>0.461</td>
</tr>
<tr>
<td>Baseline + mov</td>
<td>0.738</td>
<td>0.321</td>
<td>0.485</td>
</tr>
<tr>
<td>Baseline + pos</td>
<td>0.644</td>
<td>0.345</td>
<td>0.484</td>
</tr>
<tr>
<td>Baseline + mov + pos (ours)</td>
<td><strong>0.760</strong></td>
<td><strong>0.356</strong></td>
<td><strong>0.508</strong></td>
</tr>
</tbody>
</table>

Object-movement-direction normalization (baseline + mov) is more effective for activity with movement.
Effects of two types normalization

<table>
<thead>
<tr>
<th>Method</th>
<th>mAP (car activity)</th>
<th>mAP (car-person activity)</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.583</td>
<td>0.307</td>
<td>0.461</td>
</tr>
<tr>
<td>Baseline + mov</td>
<td>0.738</td>
<td>0.321</td>
<td>0.485</td>
</tr>
<tr>
<td>Baseline + pos</td>
<td>0.644</td>
<td>0.345</td>
<td>0.484</td>
</tr>
<tr>
<td>Baseline + mov + pos (ours)</td>
<td><strong>0.760</strong></td>
<td><strong>0.356</strong></td>
<td><strong>0.508</strong></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Method</th>
<th>nAUDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline [Sun+ TRECVID19]</td>
<td>0.589</td>
</tr>
<tr>
<td>Baseline + mov. + pos. (ours)</td>
<td>0.579</td>
</tr>
</tbody>
</table>
Conclusion

● 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