Surveillance Event Detection

NHK STRL

†Masaki Takahashi, †Masahide Naemura, †Mahito Fujii, ‡Shin'ichi Satoh

†Japan Broadcasting Corporation (NHK) Science and
Technology Research Laboratories

‡National Institutes of Informatics

Surveillance Event Detection (SED) task
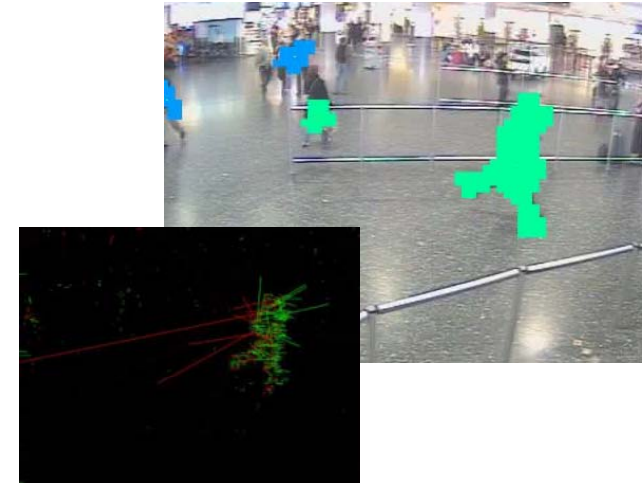
Automatic extraction of specific action sequences

Proposed method

2008

▶ Optical flow-based

- Objects were detected on the basis of direction and magnitude of optical flow

- Events were detected by using random forest classifier

2009

▶ Human region detection-based

- Human regions were detected by calculating HOG features

- Events were detected on the basis of features from a single trajectory

Proposed method

▶ Bag-of-trajectories-based

- Extracted several key-point trajectories

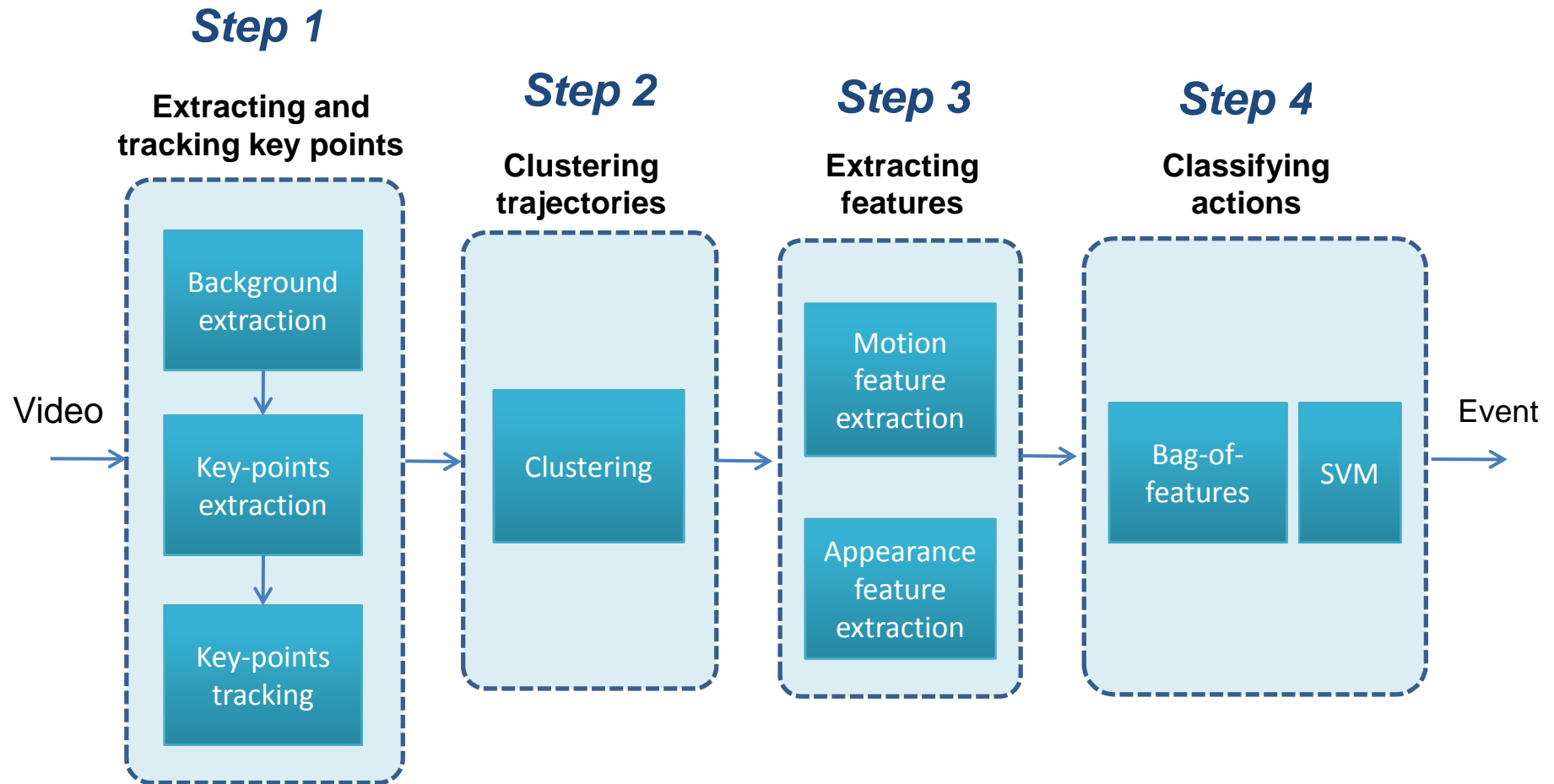- Recognized events on the basis of histogram of optical flows

⇨ Bag: entire image

▶ Enhanced bag-of-trajectories

- Clusters the human regions

⇨ Bag: a cluster region

- Considers the magnitude of optical flow in addition to its direction

Processing flow

Step 1

Extracting and tracking key points

Step 2

Clustering trajectories

Step 3

Extracting features

Step 4

Classifying actions

Video →

- Background extraction
- Key-points extraction
- Key-points tracking

- Clustering

- Motion feature extraction
- Appearance feature extraction

- Bag-of-features
- SVM

→ Event

▷ Events are recognized for each person by clustering human regions

▷ The feature is invariant of the duration of trajectories
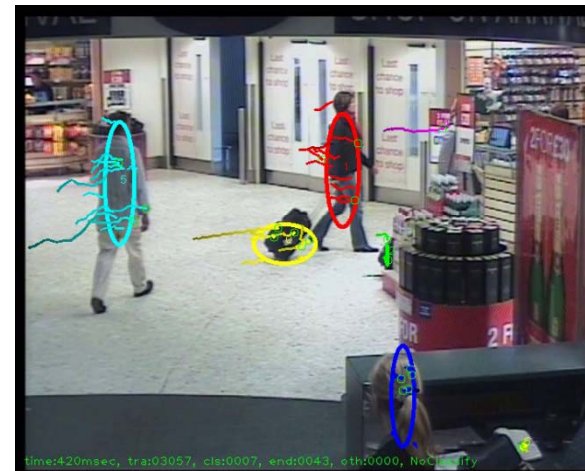
Step 1: Key-points detection & tracking

Key-point detection & tracking

Background subtraction

↓
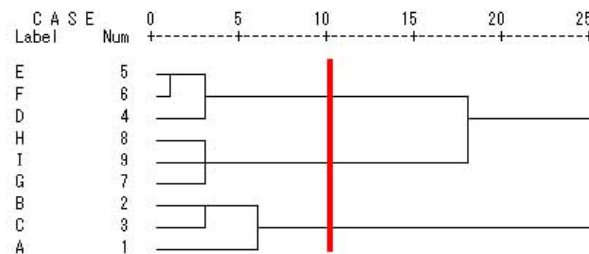
Key-points detection

↓

Key-points tracking
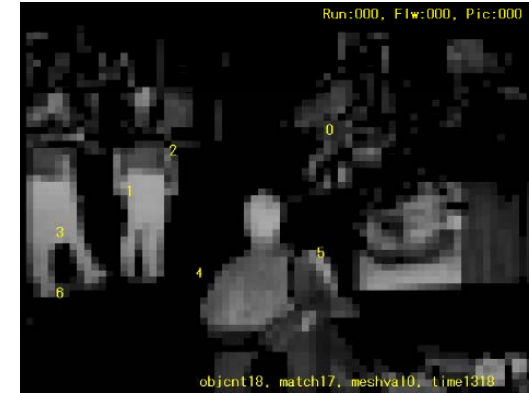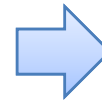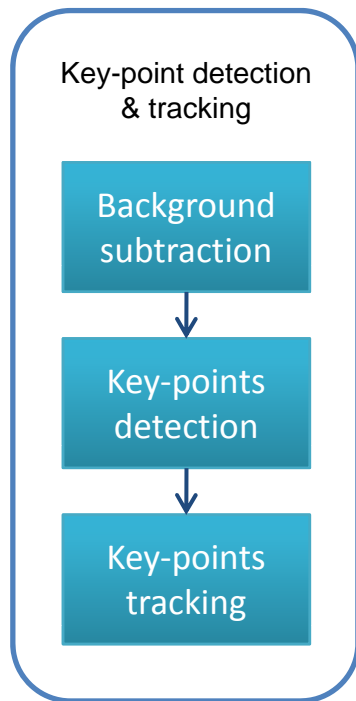
Background subtraction

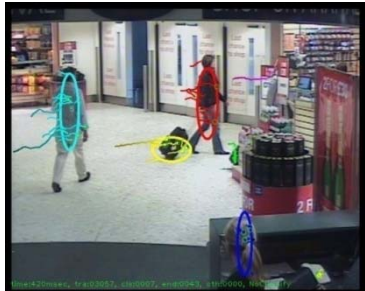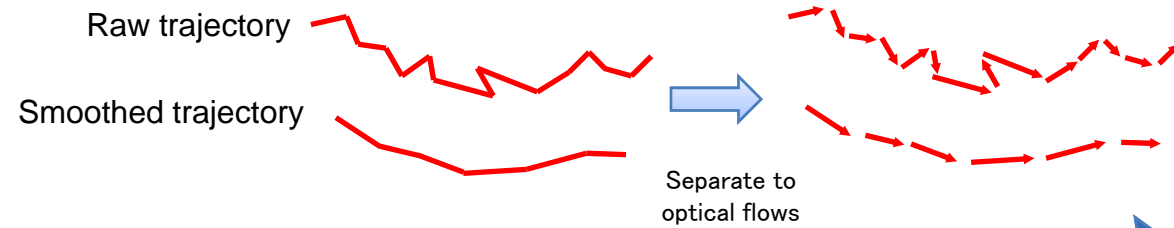Key-point trajectories

Step 2: Clustering human regions

▶ Clustering object regions on the basis of distance between
each trajectory

▷ Each cluster region is used as a bag in the bag-of-features approach

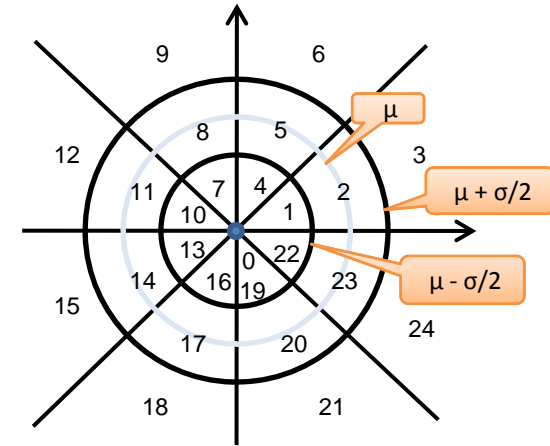▷ The threshold was decided on the basis of original annotated bounding boxes

Step 3: "Motion" feature
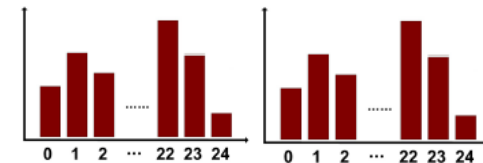
▶ Creation of trajectory histogram

Raw trajectory

Smoothed trajectory

Separate to optical flows

Calculation of mean μ、and its STD σ
of optical flows in each cluster

μ

μ + σ/2

μ - σ/2

9   6
8   5
12
11 7   4   2
10 1
13 0   22
14 16 19 23
17 20
15
18 21
24

3

Creation of histogram

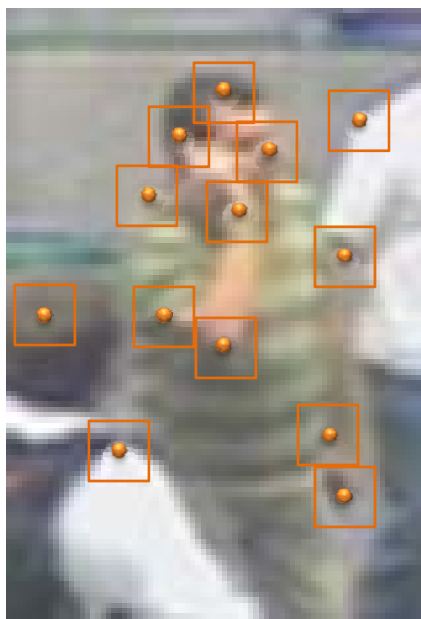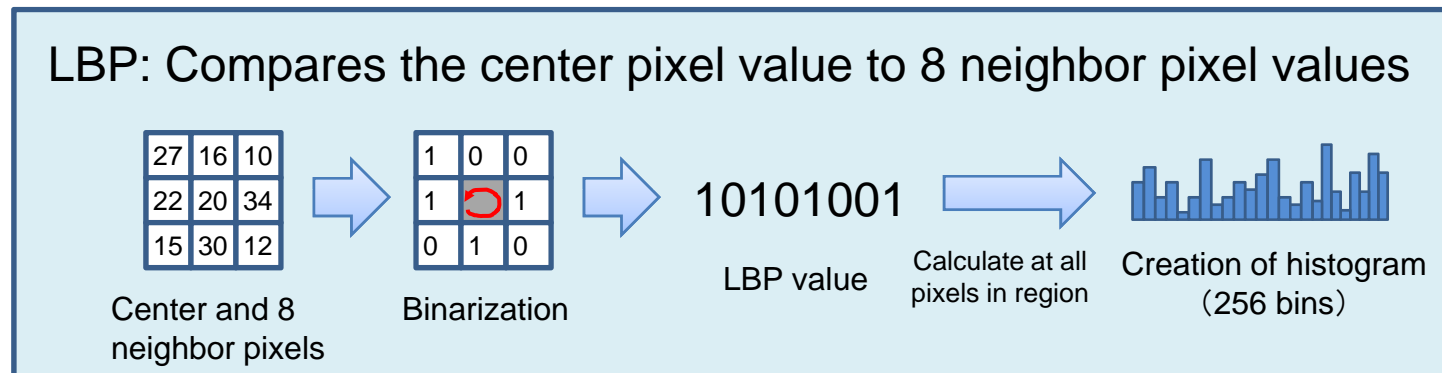Raw trajectory   Smoothed trajectory

0 1 2 ··· 22 23 24   0 1 2 ··· 22 23 24

Trajectory histogram
50 bins （25 bins × 2）

Merits of trajectory histogram

▷ Fixed feature dimension → BoF approach is possible

▷ Invariant of individual motion speed

Step 3: "Appearance" feature

▶ Creation of local binary pattern（LBP）histogram

LBP: Compares the center pixel value to 8 neighbor pixel values

27	16	10
22	20	34
15	30	12

⟹

1	0	0
1		1
0	1	0

⟹ 10101001 ⟹

Center and 8 neighbor pixels → Binarization → LBP value → Calculate at all pixels in region → Creation of histogram（256 bins）

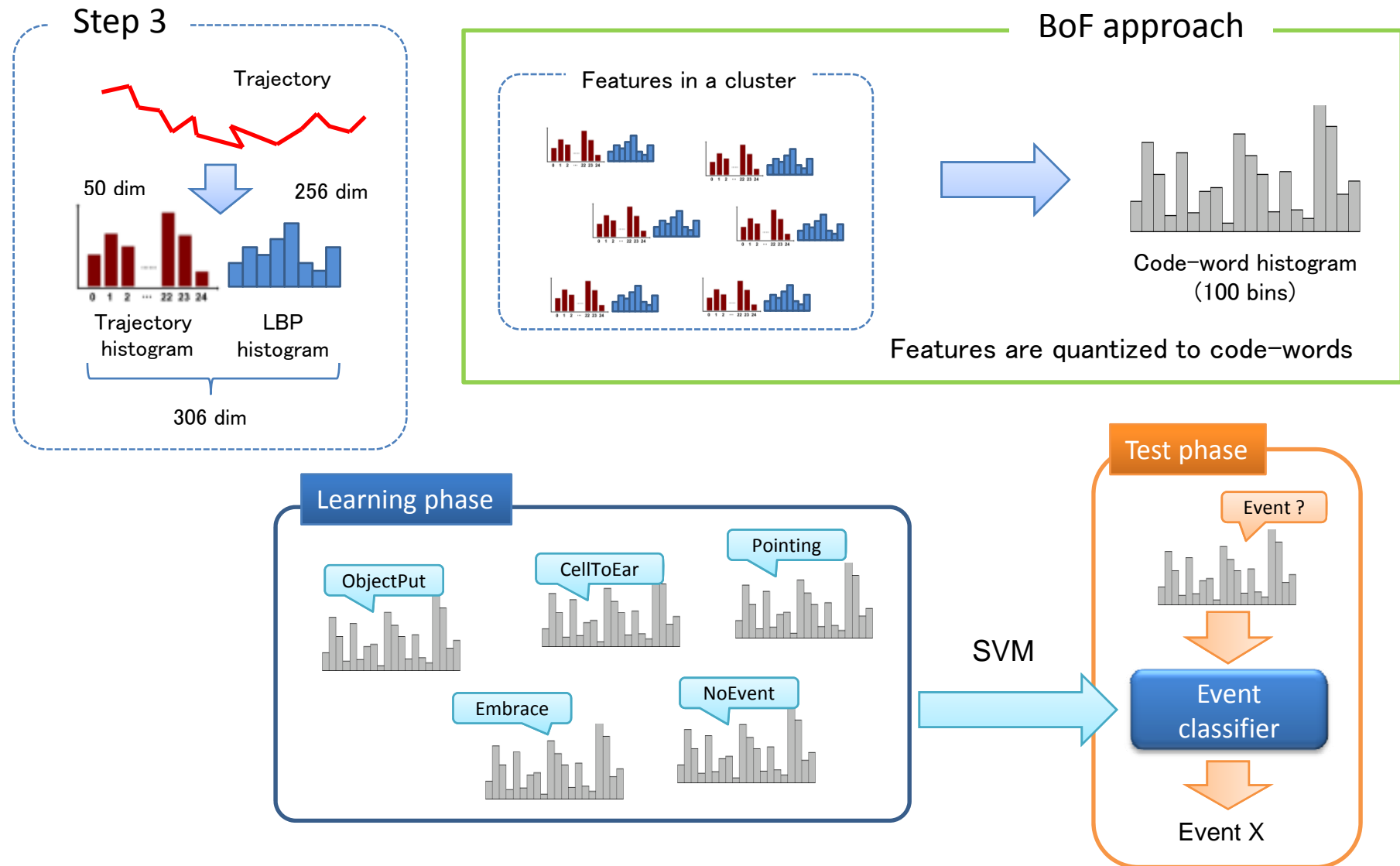Calculates LBP values at key-point centered 16 × 16 pixels

⬇

Create an LBP histogram and use it as a feature

Merits of using LBP

▷ Robust recognition by considering "appearance" feature

▷ Faster than other local image features, such as SIFT or SURF

Step 4: Event classification

▶ Bag-of-features approach + SVM

Results

▶ Evaluation results by NIST

Event	#Reference	#System	#CorDet	#FA	#Miss	DCR
CellToEar	194	165	3	162	191	1.0377
Embrace	175	835	31	804	144	1.0865
ObjectPut	621	562	10	552	611	1.1649
Pointing	1063	1278	41	1237	1022	1.3671

▷ The results for "CellToEar" were relatively stable:
the action was regular and there were few motion elements.

▷ The proposed method was not robust for events
that had significant direction variations

Experiment

▶ Comparison with human tracking-based method

Event: ObjectPut

Method	N_{ref}	N_{Sys}	N_{CorDet}	N_{FA}	N_{Miss}	R_{FA}	P_{Miss}	DCR
Human tracking-based	621	488	19	469	602	30.760	0.969	1.123
Proposed	621	1061	39	537	582	35.219	0.937	1.113

▷ The consideration of several key-point trajectories improves performance of human motion recognition

▷ Human detection and tracking is not robust with crowded video sequences

2009
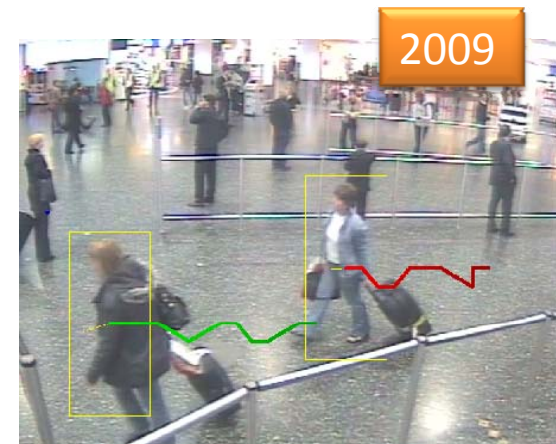
Conclusion

▶ The proposed method can robustly recognize human motion by using trajectory histograms that are invariant to individual motion speed.

▶ The extracted trajectories are clustered depending on their positions.

▶ Appearance features are considered in addition to motion features.

Future work

▷ Apply our proposed methods to other applications, such as auto tagging of motion metadata or user interfaces for TV viewing.

Thank you.