



北京邮电大学

BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

# BUPT-MCPRL@TRECVID 2016: Surveillance Event Detection

**Kaihui Zhou, Menglai Wang, Yandong Zhu,  
Yanyun Zhao, Zhicheng Zhao, Fei Su**

Multimedia Communication and Pattern Recognition Labs,  
Beijing University of Posts and Telecommunications  
**(BUPT-MCPRL)**

[zhou\\_kai\\_hui@foxmail.com](mailto:zhou_kai_hui@foxmail.com)



# Our Submission

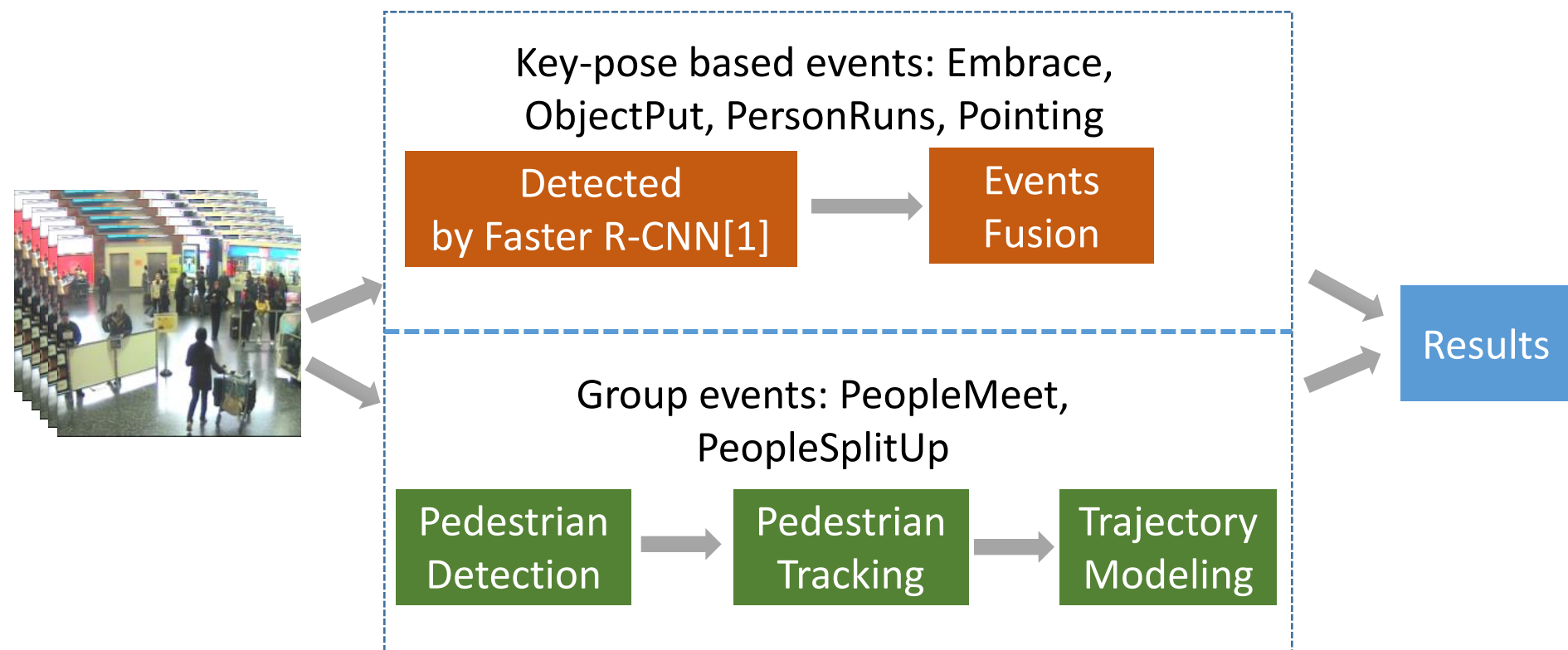
- BUPT-MCPRL 2016 Result

Event	ADCR of Our Retro-System (contrast)	Our ADCR (Interactive)	ADCR of Other Best Systems (Interactive)
Embrace	0.7207	<b>0.6622</b>	0.8448
PeopleSplitUp	0.8688	<b>0.8329</b>	0.8852
PersonRuns	0.7773	<b>0.6563</b>	0.8456
Pointing	0.9709	<b>0.9354</b>	0.9973
ObjectPut	1.0105	<b>0.9666</b>	1.0044
PeopleMeet	0.9747	<b>0.8774</b>	0.9269

# Outline

- Retrospective System Overview
- Pedestrian Detection
  - Detected by CNN
    - Embrace, Pointing, ObjectPut and PersonRuns
- Pedestrian Tracking
- Detected by Trajectory Modeling
  - PeopleMeet and PeopleSplitUp
- Conclusion

# Retrospective System Overview



# Outline

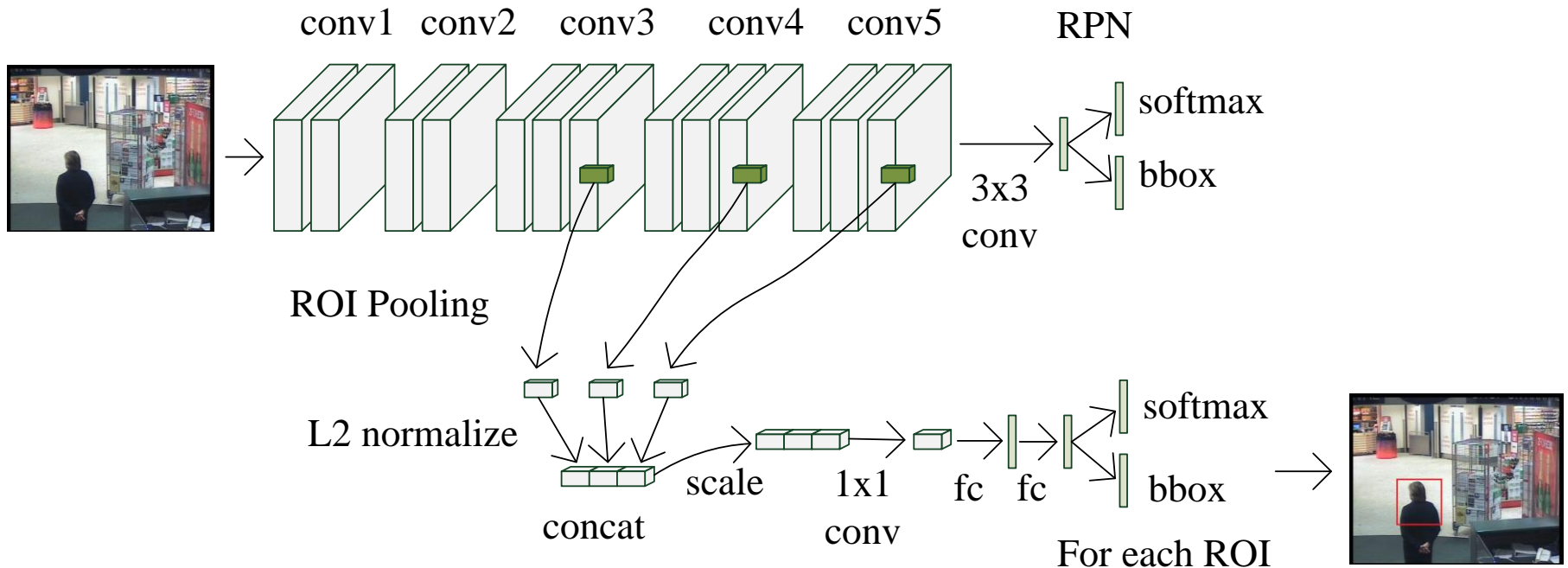
- Retrospective System Overview
- **Pedestrian Detection**
- Detected by CNN
  - Embrace, Pointing, ObjectPut and PersonRuns
- Pedestrian Tracking
- Detected by Trajectory Modeling
  - PeopleMeet and PeopleSplitUp
- Conclusion

# Pedestrian Detection

- Pedestrian Detection by Faster R-CNN
  - head-shoulder detection
    - Most pedestrian instances (about 73%) show incomplete body parts
    - More than 98% keep head and shoulder
  - Two main drawbacks
    - Conv5\_3 feature maps are of low resolution
    - Lack hard negative mining

# Pedestrian Detection

- Pedestrian Detection by Faster R-CNN
  - Skip pooling, following the work of Inside-Outside Net[2]



# Pedestrian Detection

- Pedestrian Detection by Faster R-CNN
  - Hard negative mining
    - Train a pedestrian/background classifier and detect pedestrians on the training set
    - Collect negative instances :
      - IOU overlaps  $< 0.5$  (with any ground truth)
      - and scores  $> 0.8$
    - Retrain a three-class classifier and fine-tune it from the two-class classifier

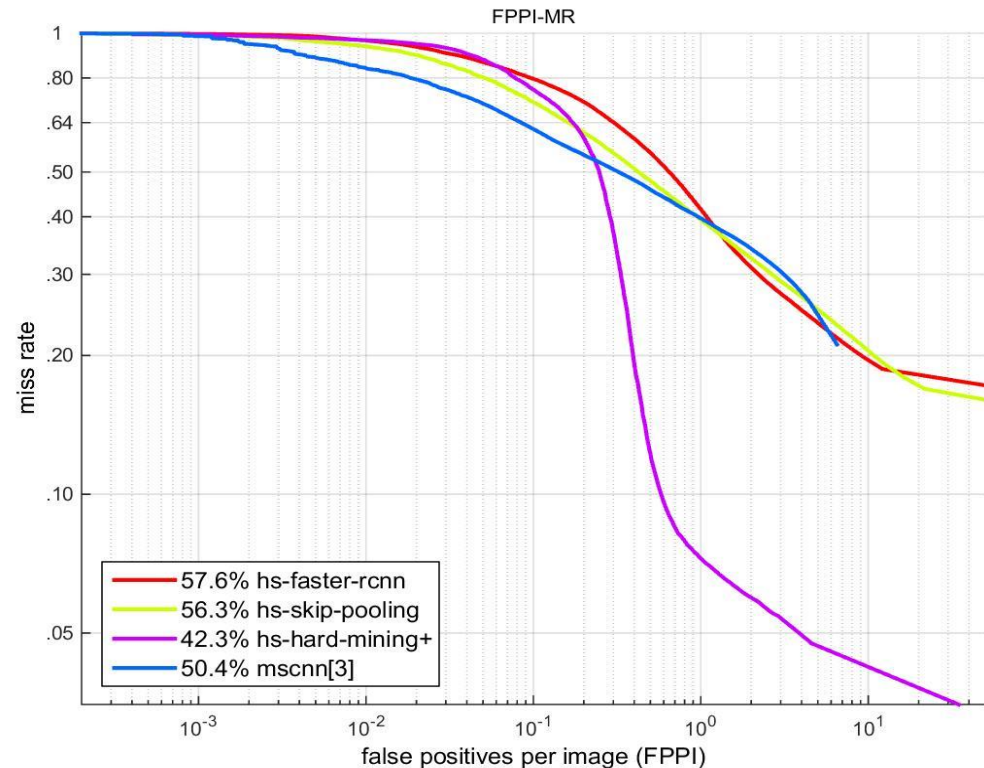


# Pedestrian Detection

- Dataset:
  - SED-PD.v2: re-annotated to be more precise
  - Randomly sampled from TrecVid08-Dataset
  - positive
    - 156,000 for training
    - 87,000 for validation
    - 257,000 for testing
  - negative
    - anything of non-positive

# Pedestrian Detection

- Results on SED-PD
  - Skip pooling, 1.3% improvement in MR
  - Hard example mining, 14% improvement in MR
  - Better results than MS-CNN (h720-ctx) [3]



[3]Cai, Zhaowei, et al. ECCV 2016

# Outline

- Retrospective System Overview
- Pedestrian Detection
- **Detected by CNN**
  - Embrace, Pointing, ObjectPut and PersonRuns
- Pedestrian Tracking
- Detected by Trajectory Modeling
  - PeopleMeet and PeopleSplitUp
- Conclusion

# Embrace, Pointing, ObjectPut and PersonRuns

- Regard four actions as the classification of static **key-poses**



**Pointing**



**Embrace**



**ObjectPut**



**PersonRuns**

# Embrace, Pointing, ObjectPut, and PersonRuns

- **Embrace, Pointing, ObjectPut**
  - Apply the original Faster R-CNN
- **PersonRuns**
  - Merge the optical flow features into original static pictures
- **Event fusion**
  - Nearest neighbor matching

# Embrace and Pointing

Embrace:

Year	ADCR	#Targ	#CorDet	#FA	#Miss
2015	0.7909	138	36	90	102
2016	0.7207	173	58	112	115

Pointing:

Year	ADCR	#Targ	#CorDet	#FA	#Miss
2015	1.0040	794	16	42	778
2016	0.9709	929	91	138	838

# ObjectPut and PersonRuns

## ObjectPut:

Year	ADCR	#Targ	#CorDet	#FA	#Miss
2015	1.0120	289	2	33	287
2016	1.0105	348	12	90	336

## PersonRuns:

Year	ADCR	#Targ	#CorDet	#FA	#Miss
2015	0.9700	50	4	87	46
2016	0.7773	63	19	158	44

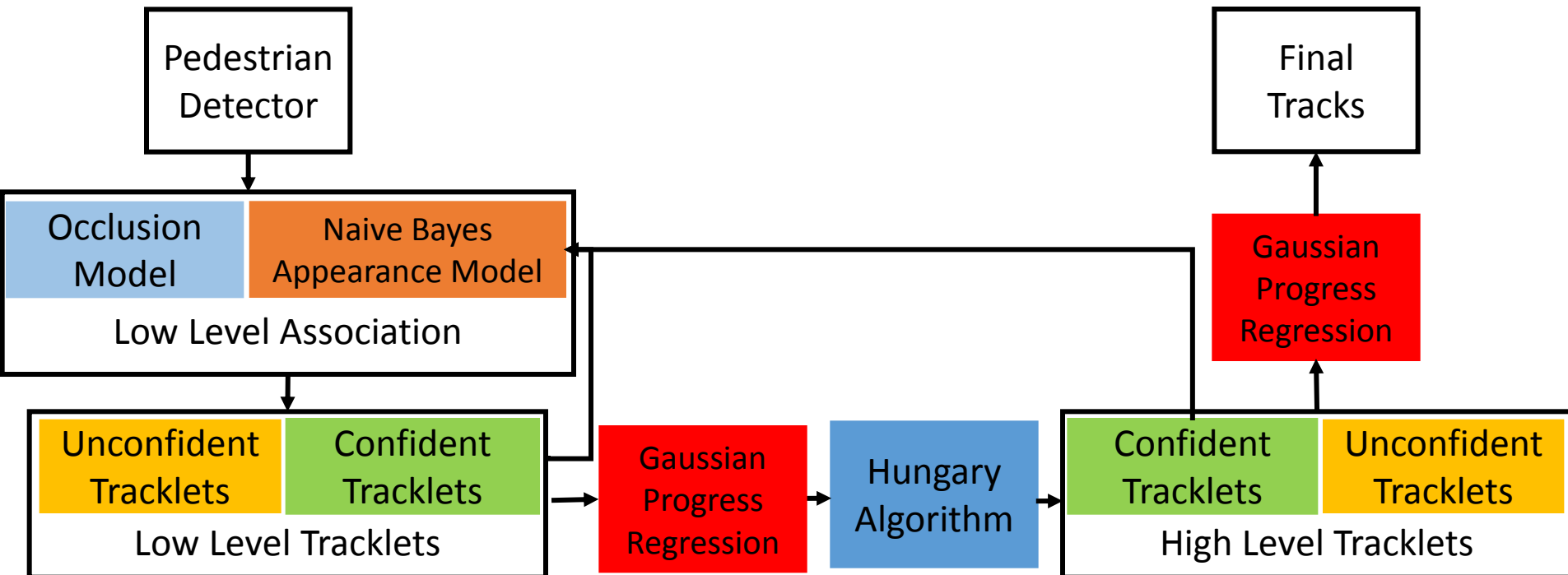
# Outline

- Retrospective System Overview
- Pedestrian Detection
  - Detected by CNN
    - Embrace, Pointing, ObjectPut and PersonRuns
- **Pedestrian Tracking**
- Detected by Trajectory Modeling
  - PeopleMeet and PeopleSplitUp
- Conclusion



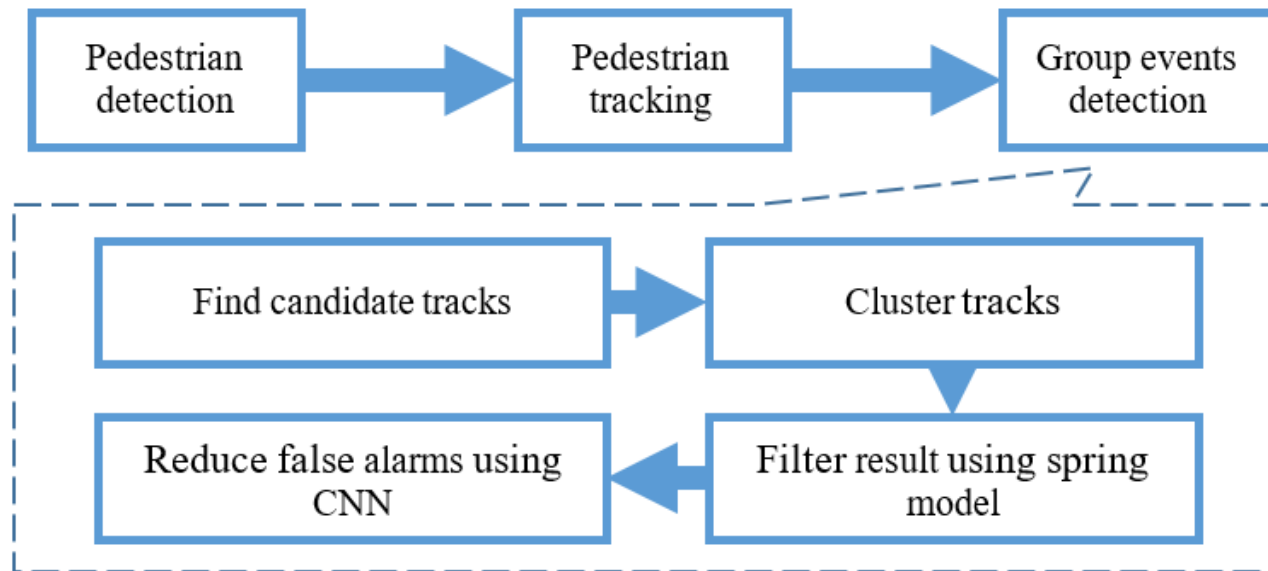
# Pedestrian Tracking

- Still followed this tracking method [Bo Yang. CVPR 2012]
  - Motion and location information are included
  - 20 times faster tracking algorithm
  - Naive Bayesian model to handle occlusion



# PeopleMeet and PeopleSplitUp

- Framework of group events detection
- Two modifications
  - Spring model
  - CNN classifier for interaction



Detection framework. Group events detection in detail is shown in the dash-line box.

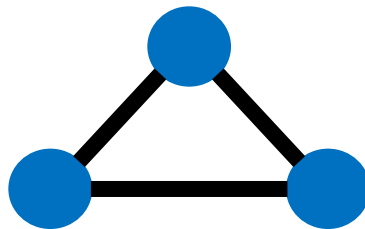
# Spring model

- Hooker's Law:  $F = kx$
- Potential energy:  $E = \frac{1}{2}kx^2$
- For a group of  $N$  pedestrians, at time  $t$  :

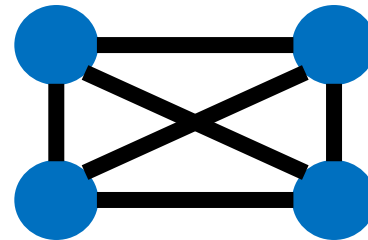
$$E(t) = \sum_{i=1}^{N_s} k[x_i(t) - x_{i0}]^2$$



$$N_p = 2 \quad N_s = 1$$



$$N_p = 3 \quad N_s = 3$$

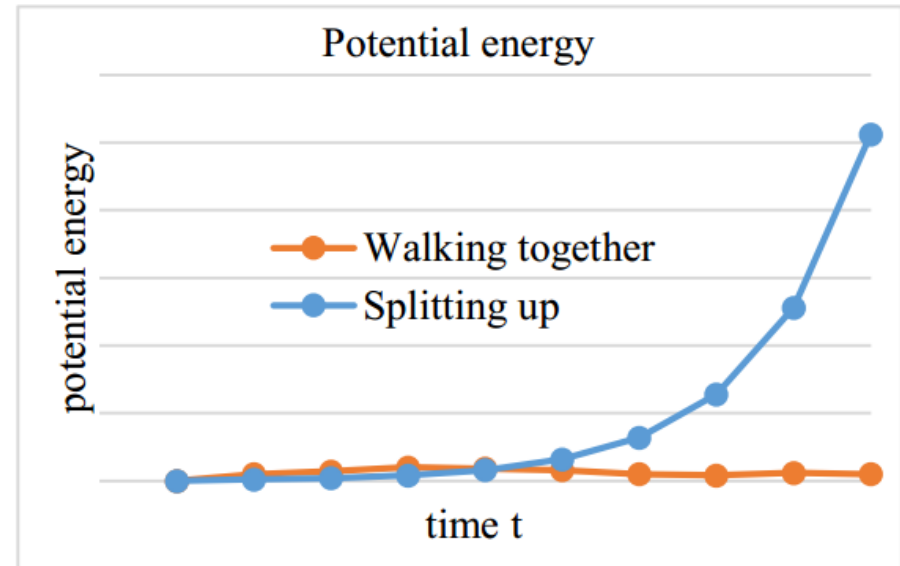
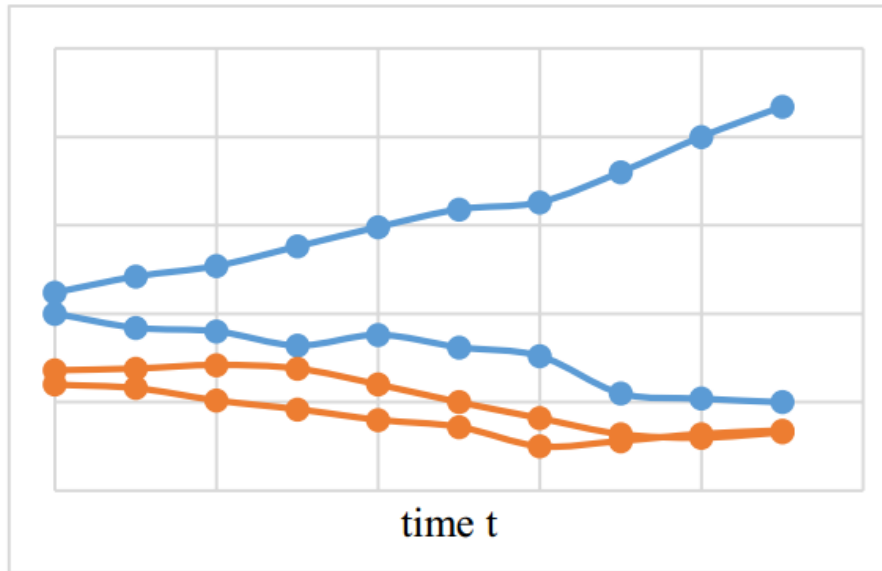


$$N_p = 4 \quad N_s = 6$$

Spring model constructed on groups of 2, 3 and 4 pedestrians, respectively

# Spring model

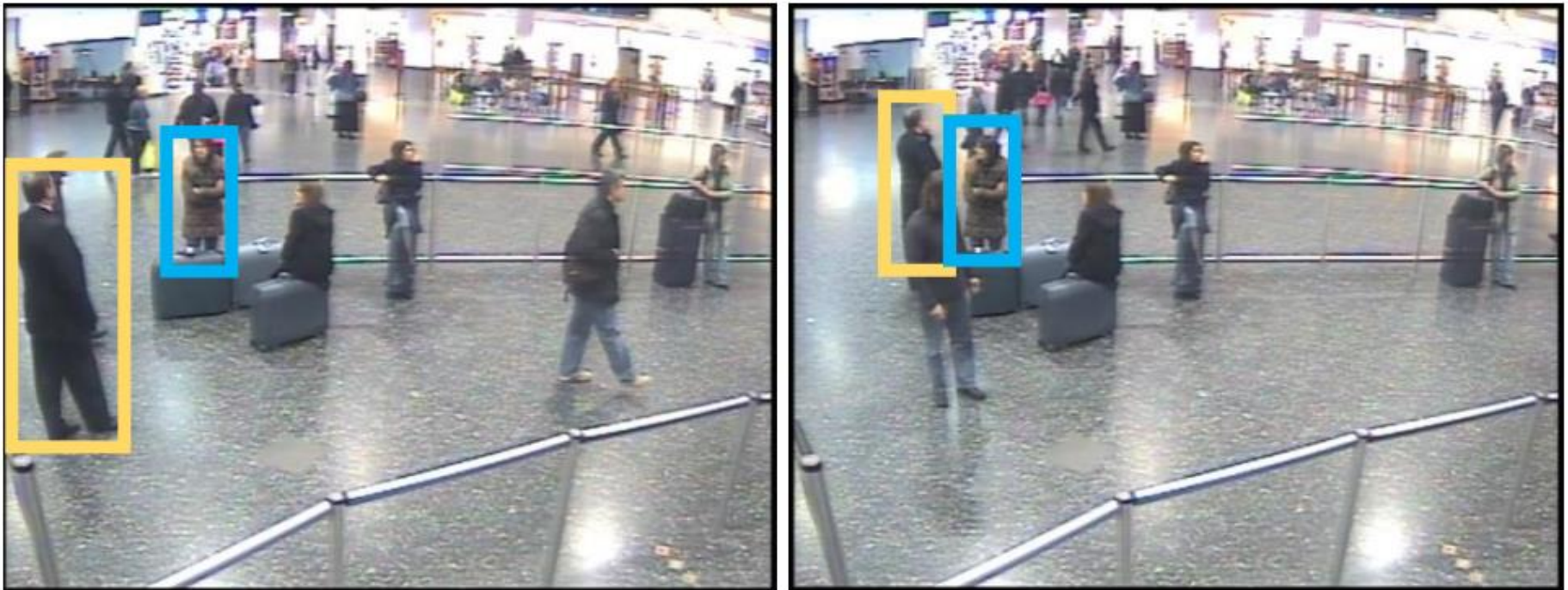
- Potential energy



The evolution of potential energy. The left figure shows the tracklets of two groups of people. The pedestrians w.r.t. blue lines are splitting up while the pedestrians w.r.t. orange lines are walking together, side by side. The right figure shows the changes of corresponding potential energy.

# PeopleMeet and PeopleSplitUp

- False alarm case
  - train a CNN to recognize the “head orientation” of pedestrians by 4 classes(front, back, left and right)



False alarm case. The man in yellow box is walking toward the woman in blue box, stopping by her side. The spring model would consider it as a meet event but there is no interaction between the man and the woman.

# PeopleMeet and PeopleSplitUp

PeopleMeet :

Year	ADCR	#Targ	#CorDet	#FA	#Miss
2015	1.0426	256	30	278	226
2016	0.9747	323	68	371	255

PeopleSplitUp:

Year	ADCR	#Targ	#CorDet	#FA	#Miss
2015	0.9387	152	24	168	128
2016	0.8688	176	47	272	129

# Conclusion

- SED-PD.v2: more accurate data make better performance.
- Faster R-CNN: FCN architecture with RoI pooling scheme allows much faster and more effective at feature extraction
- Key-pose-based event detection philosophy works.
- Spring model offers an abstract description of “PeopleMeet” and “PeopleSplitUp” events.
- In future work, CNN+RNN based algorithm could be explored.

**Thank You !**

<http://mmc.sice.bupt.cn>

**BUPT-MCPRL**

