PKU@TRECVID2009: Single-Actor and Pair-Activity Event Detection in Surveillance Video

General Coach: Wen Gao a, Xihong Wu b, Tiejun Huang a
Executive Coach: Yonghong Tian a, Yaowei Wang a, Lei Qing a
Member: Zhipeng Hu a*, Guangnan Ye b*, Guochen Jia a, Xibin Chen b, Qiong Hu c, Kaihua Jiang b

a National Engineering Laboratory for Video Technology, Peking University
b Speech and Hearing Research Center, Peking University
c Key Lab of Intel. Inf. Proc., Institute of Computing Technology, Chinese Academy of Sciences
Outline

☐ Overview
  ■ Introduction of TRECVID-ED Tasks
  ■ Summary of TRECVID-ED 2008
  ■ Our Results in TRECVID-ED 2009

☐ Our Solution in the eSur System
  ■ Background Modeling
  ■ Detection and Tracking
  ■ Event Classification
  ■ Post-processing

☐ Illustrative Results

☐ Summary
Overview of TRECVID-ED Tasks

☐ Task
  - To develop an automatic system to detect observable events in surveillance video

☐ Ten Events
  - PeopleMeet
  - PeopleSplitUp
  - Embrace
  - ElevatorNoEntry
  - PersonRun
  - CellToEar
  - ObjectPut
  - TakePicture
  - Pointing
  - OpposingFlow

☐ Challenges
  - Clutter scenes
  - Illumination variations
  - Occlusion
  - Different camera views
  - No clear event definition
The Best Results of 2008

<table>
<thead>
<tr>
<th>SITEID</th>
<th>Event</th>
<th>#Ref</th>
<th>#Sys</th>
<th>#CorDet</th>
<th>#FA</th>
<th>#Miss</th>
<th>Act.DCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFP-UIUC-NEC</td>
<td>CellToEar</td>
<td>349</td>
<td>15</td>
<td>1</td>
<td>14</td>
<td>348</td>
<td>0.999</td>
</tr>
<tr>
<td>Intuvision</td>
<td>ElevatorNoEntry</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>DCU</td>
<td>Embrace</td>
<td>401</td>
<td>36193</td>
<td>91</td>
<td>5091</td>
<td>310</td>
<td>1.271</td>
</tr>
<tr>
<td>IFP-UIUC-NEC</td>
<td>ObjectPut</td>
<td>1944</td>
<td>83</td>
<td>6</td>
<td>77</td>
<td>1938</td>
<td>1.004</td>
</tr>
<tr>
<td>Intuvision</td>
<td>OpposingFlow</td>
<td>12</td>
<td>31</td>
<td>9</td>
<td>12</td>
<td>3</td>
<td>0.251</td>
</tr>
<tr>
<td>SJTU</td>
<td>PeopleMeet</td>
<td>1182</td>
<td>25033</td>
<td>270</td>
<td>5779</td>
<td>912</td>
<td>1.337</td>
</tr>
<tr>
<td>CMU</td>
<td>PeopleSplitUp</td>
<td>671</td>
<td>42415</td>
<td>185</td>
<td>42230</td>
<td>486</td>
<td>4.856</td>
</tr>
<tr>
<td>MCG-ICT-CAS</td>
<td>PersonRuns</td>
<td>314</td>
<td>662</td>
<td>23</td>
<td>639</td>
<td>291</td>
<td>0.989</td>
</tr>
<tr>
<td>SJTU</td>
<td>Pointing</td>
<td>2316</td>
<td>1005</td>
<td>35</td>
<td>970</td>
<td>2281</td>
<td>1.080</td>
</tr>
<tr>
<td>Intuvision</td>
<td>TakePicture</td>
<td>23</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>23</td>
<td>1.000</td>
</tr>
</tbody>
</table>

- There are much rooms for improvement.
- OpposingFlow event has good detection performance.
- ElevatorNoEntry and TakePicture events are zero CorDets.
Approaches in 2008

- PeopleMeet (SJTU):
  - Camshift guided particle filter + HMM
  - Combine Head top detector and human detector
  - Camshift guided particle filter to obtain trajectory
  - HMM models to detect hidden states defined by trajectory features.

- PeopleSplitUp (CMU):
  - Key points + SVM
  - Cluster interest points into visual keywords
  - SVM classifiers to detect activities
  - Event segmentation was done in a multi-resolution framework, where all activity durations found in training were tried.

- Embrace (DCU):
  - Pedestrian tracking in 3D space
  - Detect and track pedestrians to infer the 3D location
  - Calculate the probability of person taking part in Embrace events.

- PersonRuns (ICT):
  - Data correlation + trajectory features
  - Train full-body and head-shoulder detectors using standard haar-like features
  - Adopt the data correlation method with the visual features to track objects
  - Event detection by trajectory length, location of trajectory points and speed.

- ElevatorNoEntry (INTUVISION):
  - Pedestrian detection + histogram matching
  - Haar object pedestrian detection
  - Histogram matching to find person not entering an elevator

P. Yarlagadda, et. al, INTUVISION EVENT DETECTION SYSTEM FOR TRECVID 2008

X. Yang, et al., Shanghai Jiao Tong University participation in high-level feature extraction, automatic search and surveillance event detection at TRECVID 2008

A. Hauptmann et al. INFORMEDIA @ TRECVID2008: Exploring New Frontiers
## Our Results in TRECVID-ED2009 (1)

<table>
<thead>
<tr>
<th>Event</th>
<th>#Ref</th>
<th>#Sys</th>
<th>#CorDet</th>
<th>#FA</th>
<th>#Miss</th>
<th>Act. DCR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>p-eSur_1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PeopleMeet</td>
<td>449</td>
<td>125</td>
<td>7</td>
<td>118</td>
<td>442</td>
<td>1.023</td>
</tr>
<tr>
<td>PeopleSplitUp</td>
<td>187</td>
<td>198</td>
<td>7</td>
<td>191</td>
<td>180</td>
<td>1.025</td>
</tr>
<tr>
<td>Embrace</td>
<td>175</td>
<td>80</td>
<td>1</td>
<td>79</td>
<td>174</td>
<td>1.020</td>
</tr>
<tr>
<td>ElevatorNoEntry</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0.334</td>
</tr>
<tr>
<td><strong>p-eSur_2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PeopleMeet</td>
<td>449</td>
<td>210</td>
<td>15</td>
<td>195</td>
<td>434</td>
<td>1.030</td>
</tr>
<tr>
<td>PeopleSplitUp</td>
<td>187</td>
<td>881</td>
<td>14</td>
<td>867</td>
<td>173</td>
<td>1.209</td>
</tr>
<tr>
<td>Embrace</td>
<td>175</td>
<td>164</td>
<td>3</td>
<td>161</td>
<td>172</td>
<td>1.036</td>
</tr>
<tr>
<td>PersonRuns</td>
<td>107</td>
<td>356</td>
<td>5</td>
<td>351</td>
<td>102</td>
<td>1.068</td>
</tr>
<tr>
<td><strong>p-eSur_3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PeopleMeet</td>
<td>449</td>
<td>210</td>
<td>15</td>
<td>195</td>
<td>434</td>
<td>1.030</td>
</tr>
<tr>
<td>PeopleSplitUp</td>
<td>187</td>
<td>881</td>
<td>14</td>
<td>867</td>
<td>173</td>
<td>1.209</td>
</tr>
<tr>
<td>Embrace</td>
<td>175</td>
<td>164</td>
<td>3</td>
<td>161</td>
<td>172</td>
<td>1.036</td>
</tr>
<tr>
<td>ElevatorNoEntry</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Our Results in TRECVID-ED2009 (2)

- Compared with the best results in TRECVID-ED 2008

  - Directly on the reported results in terms of Act. DCR

<table>
<thead>
<tr>
<th>Event</th>
<th>Our Best</th>
<th>Best 2008</th>
<th>Imp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PeopleMeet</td>
<td>1.023</td>
<td>1.337</td>
<td>-0.314</td>
</tr>
<tr>
<td>PeopleSplitUp</td>
<td>1.025</td>
<td>4.856</td>
<td>-3.831</td>
</tr>
<tr>
<td>Embrace</td>
<td>1.020</td>
<td>1.271</td>
<td>-0.251</td>
</tr>
<tr>
<td>ElevatorNoEntry</td>
<td>0.334</td>
<td>N/A</td>
<td>-</td>
</tr>
<tr>
<td>PersonRuns</td>
<td>1.068</td>
<td>0.989</td>
<td>+0.079</td>
</tr>
</tbody>
</table>

Note: Our results are evaluated on the ED 2009 data by 2009 DCR metric, while the 2008 best results are evaluated on the ED 2008 data by 2008 DCR metric.

- On the TRECVID-ED 2008 data in terms of 2008 Act. DCR

<table>
<thead>
<tr>
<th>Event</th>
<th>Our Best</th>
<th>Best 2008</th>
<th>Imp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PeopleMeet</td>
<td>1.245</td>
<td>1.337</td>
<td>-0.092</td>
</tr>
<tr>
<td>PeopleSplitUp</td>
<td>1.976</td>
<td>4.856</td>
<td>-2.880</td>
</tr>
<tr>
<td>Embrace</td>
<td>1.208</td>
<td>1.271</td>
<td>-0.063</td>
</tr>
<tr>
<td>ElevatorNoEntry</td>
<td>0.130</td>
<td>N/A</td>
<td>-</td>
</tr>
<tr>
<td>PersonRuns</td>
<td>1.249</td>
<td>0.989</td>
<td>+0.260</td>
</tr>
</tbody>
</table>
What are Improved?

**What?**
1. Effectively reduce the false alarms of detection
2. Obtain comparable detection accuracy, and much better results for ElevatorNoEntry

**Why?**
1. Adaptive background modeling
2. Effective human detection and tracking
3. Ensemble of one-vs.-all SVM and automata-based classifiers
4. Effective event merging and post-processing
Our Solution: Treatments for Different Event Categories

- **Pair-activity Event:**
  - One people interact with another people

- **Single-actor Event:**
  - No interaction with other people
Our eSur Framework for TRECVID-ED

- Camera Classification
- Background Subtraction
  - Body Detection
  - Head-Shoulder Detection
  - Object Tracking
- Feature Extraction
- One VS All SVM
- Automata
- Post-Processing
- Events Merging

Final System Outputs

Input: Video Frame

Training Data:
- Camera1 Training Data
- Camera2 Training Data
- Camera3 Training Data
- Camera4 Training Data
- Camera5 Training Data

Feature Extraction

Prior Knowledge

Preliminary Events

Event interval
Our Solution (1): Background Modeling

- **Mixture of Gaussian (MoG):**
  - To accurately extract the foreground while effectively decreasing detection false alarms.

- **Block-wise PCA Model:**
  - To identify which camera the video belongs to
    - Also used in the ElevatorNoEntry event detection.
  - “block” : segment each frame into blocks
  - “wise” : adaptively select the principle component for background reconstruction
MoG

- **Key Idea**
  - Randomly select 1000 frames from each camera
  - Manually label the foreground objects
  - Use EM algorithm to estimate the model

- **Results of Background Reconstruction**

- **Disadvantage: Computation time-consuming**
Block-wise PCA

- **General PCA**
  - Model a whole frame
  - **Problems**
    - high spatio-temporal computation complexity
    - high miss ratio (especially for static objects).

- **Block-wise PCA**
  - Segment a frame into blocks, and model each block respectively.
  - Lower spatio-temporal computation complexity
  - Adaptively select principle component by the MMSE to the mean background
  - Lower miss ratio and less block effect.

\[
B = \arg\min_{B_i} \| \bar{T} - B_i \|^2 \quad B_i = \phi_i \phi_i^T I
\]

where \( \bar{T} \) is the trained mean background, \( \phi_i \) is the \( i \)th principle component and \( B_i \) is the \( i \)th reconstructed background.
Comparative Results

☐ Blocking vs. No Blocking

<table>
<thead>
<tr>
<th>Method</th>
<th>No-blocking</th>
<th>Blocking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training time (for 300 frames)</td>
<td>361.332s</td>
<td>150.406s</td>
</tr>
</tbody>
</table>

* Experiment platform: Intel Xeon E5410 2.33GHz, 8G

☐ Block PCA vs. Block-wise PCA

original image | Block PCA | Block-wise PCA
Our Solution (2): Detection and Tracking

- Detection: Histogram of oriented gradients (HOG) for both whole body and head-shoulder
- Tracking: Online boosting
  - Forward and backward tracking
  - Combining color similarity to reduce drift
HOG Detector

- Fusion of Head-shoulder and Body detection
- Adjust the detector searching scales
Detection Results
Tracking Process

Frame 1.  2.  3.  4.  ...

Forward Tracking
Backward Tracking

Combined Result:
Expected Target: ○ ○ ○
Detection Result: ○ ○ ○
Canceled: ■ ■ ■
Expected Path:  →
Final Path: →
State Machine of Tracking

D : Detection existence
ND: No detection results
P : Online boosting prediction result
NH: Not human, drifting happens
H : No drifting
S : Online boosting and detection results are similar
U : Online boosting and detection results aren’t similar

Head –shoulder and Body Detection
Start
D
Start Prediction
P
Backward Tracking

ND

H

S

U

End

Trigger Another Tracking Process

Combine the Results and Trigger Another Tracking Process

Judge Prediction Probability

Judge Prediction Probability

Similarity Comparison

Similarity Comparison
Detection and Tracking Results

Detection Results

Tracking Results
Drift Reduction by Color Similarity

- **Problem**: Drifting
- **Solution**: Combine color similarity to refine tracking results

![Tracking Result without Color Similarity Comparison](image1)

![Tracking Result with Color Similarity Comparison](image2)
Our Solution (3):
Events Detection - Pair-activity

- Event Analysis using key frames
  - Key Frames: Frames characterize an event happening
  - “PeopleMeet” and “Embrace”
    - At the end of the event
  - “PeopleSplitUp”
    - At the beginning of the event
Events Detection - Pair-activity

- **Relational Features**

Three matrices $DT$, $MD$, $CT$ in the $n$th frame:

- $DT_{ij} = dist(obj_i, obj_j)$
- $MD_{ij} = relCode(obj_i, obj_j)$
- $CT_{ij} = cotime(obj_i, obj_j)$

Feature vector between objects $i$ and $j$ in the $n$th frame:

$F_n(i,j) = \{ DT_n(i,j), MD_n(i,j), CT_n(i,j) \}$

**Distance**

$$dist(obj_1, obj_2) = \| pos_{obj1} - pos_{obj2} \|$$

**Motion Direction Correlation**

$$relCode(obj1, obj2) = \begin{cases} 
0 & \text{if } |\theta_1 - \theta_2| < \alpha \\
1 & \text{if } |\theta_1 - \theta_2| < \beta \\
2 & \text{otherwise}
\end{cases}$$

where $\theta_1$ and $\theta_2$ are the direction angles of the two persons.

$cotime(obj_1, obj_2) = \min(t_{end}(p1), t_{end}(p2)) - \max(t_{start}(p1), t_{start}(p2))$
Events Detection - Single-Actor

- PersonRuns
  - Persons with higher velocity than others
  - Motion direction consistency

- ElevatorNoEntry
  - Elevator state detection
    - Keep close
    - Opening
    - Keep open
    - Closing
  - People state detection
    - PeopleExistence vs. No People
    - Enter, Leave and Waiting
Events Classification Framework

- Feature Extracting
  - Embrace Classifier
  - PeopleMeet Classifier
  - PeopleSplitUp Classifier
  - PersonRuns Classifier

- Event Classifier
  - Pair-Activity Event Classifier
    - Embrace Classifier
    - PeopleMeet Classifier
    - PeopleSplitUp Classifier
    - PersonRuns Classifier

- Key frames
- Event Identifying
  - Backwards Search
  - Forwards Search

- Detected Embrace
- Detected PeopleMeet
- Detected PeopleSplitUp
- Detected PersonRuns

- Event Merging
- Post-processing

- Preliminary Events
Classifiers Evaluation

- Single-level SVM classifier
  - Use one-vs-all mode, train a binary classifier for each event

- Hierarchical SVM classifier
  - Hierarchically tie events and train different level classifiers

- Multi-Kernel Learning (MKL) classifier
  - Multiple kernels (RBF, linear, and poly kernels) are combined to enhance classifier performance

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Single-level classifier</th>
<th>MKL classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>PeopleMeet</td>
<td>0.744</td>
<td>0.890</td>
</tr>
<tr>
<td>PeopleSplitUp</td>
<td>0.255</td>
<td>0.320</td>
</tr>
<tr>
<td>Embrace</td>
<td>0.418</td>
<td>0.530</td>
</tr>
<tr>
<td>PersonRuns</td>
<td>0.813</td>
<td>0.590</td>
</tr>
<tr>
<td>PersonRuns</td>
<td>0.703</td>
<td></td>
</tr>
<tr>
<td>PersonRuns</td>
<td>0.728</td>
<td></td>
</tr>
<tr>
<td>PersonRuns</td>
<td>0.815</td>
<td></td>
</tr>
<tr>
<td>PersonRuns</td>
<td>0.700</td>
<td></td>
</tr>
</tbody>
</table>

- Evaluated by NDCR

- Sample 10 hours data from TREVID-ED 2008 corpus
- Use detection and tracking results
- Single-level classifier is less time-consuming
- Manually label participating objects of each event
State Machine of ElevatorNoEntry Detection

**Elevator State Detection**
- Detect foreground in elevator regions
- No foreground in elevator regions

**Keep Close**
- Closing
- Keep Close

**Keep Open**
- Opening
- Keep Open

**Calculate foreground area of the frame**

**People State Detection**
- Start
  - No People
    - Foreground area is less than some threshold
    - Elevator Closing and no foreground on scene bounds
    - Elevator Opening
  - Elevator Opening
    - No Event

- Elevator NoEntry
  - Elevator closing and waiting state when opening

- Waiting
  - People Existence
    - Foreground area exceeds some threshold
    - Elevator Closing and detect foreground on scene bounds
  - Elevator Opening
  - No Event
Results without Post-processing

Data: 80 hours video from TRECVID-ED 2008

<table>
<thead>
<tr>
<th>Part of 2008 data</th>
<th>#Ref</th>
<th>#Sys</th>
<th>#Cor Det</th>
<th>#FA</th>
<th>#Miss</th>
<th>Act. DCR</th>
<th>Act. NDCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PeopleMeet</td>
<td>796</td>
<td>1342</td>
<td>60</td>
<td>1282</td>
<td>736</td>
<td>433.949</td>
<td>1.245</td>
</tr>
<tr>
<td>People SplitUp</td>
<td>924</td>
<td>9505</td>
<td>176</td>
<td>9329</td>
<td>748</td>
<td>557.035</td>
<td>1.976</td>
</tr>
<tr>
<td>Embrace</td>
<td>279</td>
<td>1831</td>
<td>26</td>
<td>1805</td>
<td>253</td>
<td>422.891</td>
<td>1.208</td>
</tr>
<tr>
<td>Person Runs</td>
<td>200</td>
<td>2731</td>
<td>18</td>
<td>2713</td>
<td>182</td>
<td>431.825</td>
<td>1.249</td>
</tr>
</tbody>
</table>

Too many false alarms!
Our Solution (4): Post-processing

- PeopleMeet and Embrace

  - Problem: False alarms shown as below
  - Solution:

    Final distance between the two persons must be less than some threshold.
Post-processing

- PeopleSplitUp
  - Problem: False alarms shown below
  - Solution:
    1. Original distance between the two persons must be less than some threshold
    2. The two persons should not have the same motion direction
Results in TRECVID-ED 2009 (1)

**EVENT**: ElevatorNoEntry

<table>
<thead>
<tr>
<th>Analysis Report</th>
<th>#Ref</th>
<th>#Sys</th>
<th>#CorDet</th>
<th>#FA</th>
<th>#Miss</th>
<th>Act. RFA</th>
<th>Act. PMiss</th>
<th>Act. DCR</th>
<th>Min RFA</th>
<th>Min PMiss</th>
<th>Min DCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUPT-MCPRL_6 / p-baseline_6</td>
<td>3</td>
<td>23</td>
<td>2</td>
<td>21</td>
<td>1</td>
<td>1.377</td>
<td>0.333</td>
<td>0.340</td>
<td>1.377</td>
<td>0.333</td>
<td>0.340</td>
</tr>
<tr>
<td>BUPT-PRIS_1 / p-baseline_1</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.066</td>
<td>0.667</td>
<td>0.667</td>
<td>0.066</td>
<td>0.667</td>
<td>0.667</td>
</tr>
<tr>
<td>CMU_3 / p-VCUBE_1</td>
<td>3</td>
<td>1041</td>
<td>31038</td>
<td>0</td>
<td>68.078</td>
<td>0.000</td>
<td>0.340</td>
<td>7.739</td>
<td>0.000</td>
<td>0.039</td>
<td></td>
</tr>
<tr>
<td>PKU-IDM_4 / p-eSur_1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0.131</td>
<td>0.333</td>
<td>0.334</td>
<td>0.066</td>
<td>0.333</td>
<td>0.334</td>
</tr>
<tr>
<td>PKU-IDM_4 / p-eSur_3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>SJTU_3 / p-baseline_1</td>
<td>3</td>
<td>28</td>
<td>2</td>
<td>26</td>
<td>1</td>
<td>1.705</td>
<td>0.333</td>
<td>0.342</td>
<td>1.640</td>
<td>0.333</td>
<td>0.342</td>
</tr>
<tr>
<td>Toshiba_1 / p-cohog_1</td>
<td>3</td>
<td>90</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.066</td>
<td>0.333</td>
<td>0.334</td>
<td>0.656</td>
<td>0.000</td>
<td>0.003</td>
</tr>
</tbody>
</table>
## Results in TRECVID-ED 2009 (2)

### EVENT: PeopleMeet

<table>
<thead>
<tr>
<th>Analysis Report</th>
<th>#Ref</th>
<th>#Sys</th>
<th>#CorDet</th>
<th>#FA</th>
<th>#Miss</th>
<th>Act. RFA</th>
<th>Act. PMiss</th>
<th>Act. DCR</th>
<th>Min RFA</th>
<th>Min PMiss</th>
<th>Min DCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMU_3 / p-VCUBE_1</td>
<td>449</td>
<td>2130</td>
<td>58</td>
<td>2072</td>
<td>391</td>
<td>135.894</td>
<td>0.871</td>
<td>1.550</td>
<td>36.466</td>
<td>0.998</td>
<td>1.180</td>
</tr>
<tr>
<td>NHKSTRL_2 / p-NHK-SYS1_1</td>
<td>449</td>
<td>991</td>
<td>55</td>
<td>905</td>
<td>394</td>
<td>59.355</td>
<td>0.877</td>
<td>1.174</td>
<td>1.508</td>
<td>0.991</td>
<td>0.999</td>
</tr>
<tr>
<td>PKU-IDM_4 / p-eSur_1</td>
<td>449</td>
<td>125</td>
<td>7</td>
<td>118</td>
<td>442</td>
<td>7.739</td>
<td>0.984</td>
<td>1.023</td>
<td>1.705</td>
<td>0.991</td>
<td>1.000</td>
</tr>
<tr>
<td>PKU-IDM_4 / p-eSur_2</td>
<td>449</td>
<td>210</td>
<td>15</td>
<td>195</td>
<td>434</td>
<td>12.789</td>
<td>0.967</td>
<td>1.030</td>
<td>0.000</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>PKU-IDM_4 / p-eSur_3</td>
<td>449</td>
<td>210</td>
<td>15</td>
<td>195</td>
<td>434</td>
<td>12.789</td>
<td>0.967</td>
<td>1.030</td>
<td>0.000</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>SJTU_3 / p-baseline_1</td>
<td>449</td>
<td>19739</td>
<td>108</td>
<td>7706</td>
<td>341</td>
<td>505.404</td>
<td>0.759</td>
<td>3.287</td>
<td>1.443</td>
<td>0.996</td>
<td>1.003</td>
</tr>
<tr>
<td>TITGT_1 / c-EVAL_1</td>
<td>449</td>
<td>14884</td>
<td>36214522</td>
<td>87</td>
<td>952.436</td>
<td>0.194</td>
<td>4.956</td>
<td>952.436</td>
<td>0.194</td>
<td>4.956</td>
<td></td>
</tr>
<tr>
<td>TITGT_1 / p-EVAL_1</td>
<td>449</td>
<td>14161</td>
<td>35413807</td>
<td>95</td>
<td>905.542</td>
<td>0.212</td>
<td>4.739</td>
<td>905.542</td>
<td>0.212</td>
<td>4.739</td>
<td></td>
</tr>
</tbody>
</table>
### EVENT: PeopleSplitUp

<table>
<thead>
<tr>
<th>Analysis Report</th>
<th>#Ref</th>
<th>#Sys</th>
<th>#Dett</th>
<th>#FA</th>
<th>#Miss</th>
<th>Act. RFA</th>
<th>Act. PMiss</th>
<th>Act. DCR</th>
<th>Min RFA</th>
<th>Min PMiss</th>
<th>Min DCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMU_3 / p-VCUBE_1</td>
<td>187</td>
<td>10184</td>
<td>28</td>
<td>10156</td>
<td>159</td>
<td>666.088</td>
<td>0.850</td>
<td>4.181</td>
<td>0.721</td>
<td>0.995</td>
<td>0.998</td>
</tr>
<tr>
<td>PKU-IDM_4 / p-eSur_1</td>
<td>187</td>
<td>198</td>
<td>7</td>
<td>191</td>
<td>180</td>
<td>12.527</td>
<td>0.963</td>
<td>1.025</td>
<td>0.525</td>
<td>0.995</td>
<td>0.997</td>
</tr>
<tr>
<td>PKU-IDM_4 / p-eSur_2</td>
<td>187</td>
<td>881</td>
<td>14</td>
<td>867</td>
<td>173</td>
<td>56.863</td>
<td>0.925</td>
<td>1.209</td>
<td>0.066</td>
<td>0.995</td>
<td>0.995</td>
</tr>
<tr>
<td>PKU-IDM_4 / p-eSur_3</td>
<td>187</td>
<td>881</td>
<td>14</td>
<td>867</td>
<td>173</td>
<td>56.863</td>
<td>0.925</td>
<td>1.209</td>
<td>0.066</td>
<td>0.995</td>
<td>0.995</td>
</tr>
<tr>
<td>SJTU_3 / p-baseline_1</td>
<td>187</td>
<td>22877</td>
<td>66</td>
<td>11690</td>
<td>121</td>
<td>766.697</td>
<td>0.647</td>
<td>4.481</td>
<td>1.705</td>
<td>0.984</td>
<td>0.993</td>
</tr>
<tr>
<td>TITGT_1 / c-EVAL_1</td>
<td>187</td>
<td>15007</td>
<td>186</td>
<td>14821</td>
<td>1</td>
<td>972.046</td>
<td>0.005</td>
<td>4.866</td>
<td>972.046</td>
<td>0.005</td>
<td>4.866</td>
</tr>
<tr>
<td>TITGT_1 / p-EVAL_1</td>
<td>187</td>
<td>14239</td>
<td>184</td>
<td>14055</td>
<td>3</td>
<td>921.807</td>
<td>0.016</td>
<td>4.625</td>
<td>921.807</td>
<td>0.016</td>
<td>4.625</td>
</tr>
</tbody>
</table>
Results in TRECVID-ED 2009 (4)

**EVENT:** Embrace

<table>
<thead>
<tr>
<th>Analysis Report</th>
<th>#Ref</th>
<th>#Sys</th>
<th>#CorDet</th>
<th>#FA</th>
<th>#Miss</th>
<th>Act. RFA</th>
<th>Act. PMiss</th>
<th>Act. DCR</th>
<th>Min RFA</th>
<th>Min PMiss</th>
<th>Min DCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMU_3 / p-VCUBE_1</td>
<td>175</td>
<td>20080</td>
<td>146</td>
<td>19934</td>
<td>29</td>
<td>1307.38</td>
<td>0.166</td>
<td>6.703</td>
<td>1.377</td>
<td>0.989</td>
<td>0.996</td>
</tr>
<tr>
<td>NEC-UIUC_2 / c-none_1</td>
<td>175</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>175</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>PKU-IDM_4 / p-eSur_1</td>
<td>175</td>
<td>80</td>
<td>1</td>
<td>79</td>
<td>174</td>
<td>5.181</td>
<td>0.994</td>
<td>1.020</td>
<td>3.870</td>
<td>0.994</td>
<td>1.014</td>
</tr>
<tr>
<td>PKU-IDM_4 / p-eSur_2</td>
<td>175</td>
<td>164</td>
<td>3</td>
<td>161</td>
<td>172</td>
<td>10.559</td>
<td>0.983</td>
<td>1.036</td>
<td>1.312</td>
<td>0.994</td>
<td>1.001</td>
</tr>
<tr>
<td>PKU-IDM_4 / p-eSur_3</td>
<td>175</td>
<td>164</td>
<td>3</td>
<td>161</td>
<td>172</td>
<td>10.559</td>
<td>0.983</td>
<td>1.036</td>
<td>1.312</td>
<td>0.994</td>
<td>1.001</td>
</tr>
<tr>
<td>SFU_1 / p-match_1</td>
<td>175</td>
<td>6712</td>
<td>28</td>
<td>650</td>
<td>147</td>
<td>42.631</td>
<td>0.840</td>
<td>1.053</td>
<td>1.968</td>
<td>0.989</td>
<td>0.998</td>
</tr>
<tr>
<td>SJTU_3 / p-baseline_1</td>
<td>175</td>
<td>14189</td>
<td>64</td>
<td>1919</td>
<td>111</td>
<td>125.859</td>
<td>0.634</td>
<td>1.264</td>
<td>0.328</td>
<td>0.994</td>
<td>0.996</td>
</tr>
</tbody>
</table>
### EVENT: PersonRuns

<table>
<thead>
<tr>
<th>Analysis Report</th>
<th>#Ref</th>
<th>#Sys</th>
<th>#CorDet</th>
<th>#FA</th>
<th>#Miss</th>
<th>Act. RFA</th>
<th>PMiss</th>
<th>RFA</th>
<th>DCR</th>
<th>Min RFA</th>
<th>PMiss</th>
<th>DCR</th>
<th>Min RFA</th>
<th>PMiss</th>
<th>DCR</th>
<th>Min RFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUPT-MCPRL_6 / p-baseline_6</td>
<td>107</td>
<td>25275</td>
<td>78</td>
<td>25197</td>
<td>29</td>
<td>1652.563</td>
<td>0.271</td>
<td>8.534</td>
<td>1652.563</td>
<td>0.271</td>
<td>8.534</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BUPT-PRIS_1 / p-baseline_1</td>
<td>107</td>
<td>39</td>
<td>2</td>
<td>14</td>
<td>105</td>
<td>0.918</td>
<td>0.981</td>
<td>0.986</td>
<td>0.656</td>
<td>0.981</td>
<td>0.985</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMU_3 / p-VCUBE_1</td>
<td>107</td>
<td>23721</td>
<td>87</td>
<td>23634</td>
<td>20</td>
<td>1550.053</td>
<td>0.187</td>
<td>7.937</td>
<td>2.427</td>
<td>0.991</td>
<td>1.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEC-UIUC_2 / c-none_1</td>
<td>107</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>107</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEC-UIUC_2 / p-UI_1</td>
<td>107</td>
<td>157</td>
<td>1</td>
<td>38</td>
<td>106</td>
<td>2.492</td>
<td>0.991</td>
<td>1.003</td>
<td>1.180</td>
<td>0.991</td>
<td>0.997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHKSTRL_2 / p-NHKSYS1_1</td>
<td>107</td>
<td>468</td>
<td>15</td>
<td>339</td>
<td>92</td>
<td>22.234</td>
<td>0.860</td>
<td>0.971</td>
<td>21.053</td>
<td>0.860</td>
<td>0.965</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PKU-IDM_4 / p-eSur_2</td>
<td>107</td>
<td>356</td>
<td>5</td>
<td>351</td>
<td>102</td>
<td>23.021</td>
<td>0.953</td>
<td>1.068</td>
<td>3.673</td>
<td>0.981</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFU_1 / p-match_1</td>
<td>107</td>
<td>30948</td>
<td>22</td>
<td>3078</td>
<td>85</td>
<td>201.873</td>
<td>0.794</td>
<td>1.804</td>
<td>0.984</td>
<td>0.981</td>
<td>0.986</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SJTU_3 / p-baseline_1</td>
<td>107</td>
<td>2217</td>
<td>19</td>
<td>1228</td>
<td>88</td>
<td>80.539</td>
<td>0.822</td>
<td>1.225</td>
<td>21.447</td>
<td>0.981</td>
<td>1.089</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TITGT_1 / c-EVAL_1</td>
<td>107</td>
<td>11062</td>
<td>70</td>
<td>10992</td>
<td>37</td>
<td>720.918</td>
<td>0.346</td>
<td>3.950</td>
<td>720.918</td>
<td>0.346</td>
<td>3.950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TITGT_1 / p-EVAL_1</td>
<td>107</td>
<td>11019</td>
<td>70</td>
<td>10949</td>
<td>37</td>
<td>718.098</td>
<td>0.346</td>
<td>3.936</td>
<td>718.098</td>
<td>0.346</td>
<td>3.936</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toshiba_1 / p-cohog_1</td>
<td>107</td>
<td>8380</td>
<td>1</td>
<td>176</td>
<td>106</td>
<td>11.543</td>
<td>0.991</td>
<td>1.048</td>
<td>0.262</td>
<td>0.991</td>
<td>0.992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UAM_1 / p-baseline_1</td>
<td>107</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>107</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Illustrative Results

All results are obtained from TRECVID-ED 2008 (except ElevatorNoEntry) according to the ground truth.
Our participation in TRECVID-ED 2009

Events
- PeopleMeet
- PeopleSplitUp
- Embrace
- ElevatorNoEntry
- PersonRun
- CellToEar
- ObjectPut
- TakePicture
- Pointing
- OpposingFlow

Submitted 5 event detection results
4 of them obtain significantly improvements over the best results of TRECVID-ED 2008

Three-fold contributions:
- Effective strategies for adaptive background modeling, human detection and tracking
- An ensemble approach of one-vs.-all SVM and automata-based classifiers for both single-actor and pair-activity events
- Post-processing to reduce the false alarm
Summary

Future Work

- Better human detection and tracking in crowd scenes
- Better discriminative features such as temporally integrated spatial response (TISR) descriptor [Zhu, MM09]
- More effective event classification models, such as MKL and sequence learning
Thanks!

yhtian@pku.edu.cn

Our Team
(Main members in the first row)
Our Solution:
The eSur System: An e-Sir for Event Surveillance

Towards an integrated system for analyzing *archived* and *real-time* surveillan...
Basic Idea of Block-wise PCA

Select the best reconstructed background according to the MMSE to the mean background.

\[ B = \arg \min_{B_i} \| I - \sum_{i=1}^{N} I_{\text{train}}(i) \|_2 \]

Project the frame on each principle component to reconstruct background respectively.

\[ I = \frac{1}{N} \sum_{i=1}^{N} I_{\text{train}}(i) \]

Train the Mean Background

Frame Blocking

Select the best reconstructed background according to the MMSE to the mean background.

First Principle Component \( \phi_1 \)

Second Principle Component \( \phi_2 \)

Third Principle Component \( \phi_3 \)
Results of Block-wise PCA

- The background subtraction results using block-wise PCA
Detection

- HOG-based feature for both human and HS detection
  - AdaBoost for feature selection

- Cascaded structure
  - Different weak classifier for each layer for simplicity, but not SVM as in [Zhu, CVPR06]

**HOG Feature**  
(Dalal, CVPR05)

[Dalal, CVPR05, Histograms of Oriented Gradients for Human Detection]  
[Zhu, CVPR06, Fast Human Detection Using a Cascade of Histograms of Oriented Gradients]
Final Tracking Result
Illustrative Example – False Alarm

PersonRuns

ElevatorNoEntry

PeopleMeet

Embrace

PeopleSplitUp
Illustrative Example – Miss

PersonRuns

ElevatorNoEntry

PeopleMeet

Embrace

PeopleSplitUp
Comparative Results

☐ Block PCA vs. Block-wise PCA

original image

Block PCA

Block-wise PCA