数字视频编解码技术国家工程实验室
National Engineering Laboratory for Video Technology

PKU@TRECVID-ED2010: Pair-wise 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 [c]
Member: Kaihua Jiang [b], Zhipeng Hu [a], Zhongwei Chen [c], Guochen Jia [a], Ten Xu [a], Qiong Hu [c],
Qiong Hu [c], Guangcheng Zhang [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

NELVT

Outline

- Overview
  - Tasks This Year
  - Our Results This Year
- Our eSur System for ED 2010
  - Background Modeling
  - Detection and Tracking
  - Event Detection
- Summary

Tasks This Year

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

- Events in 2009
  - PeopleMeet
  - PeopleSplitUp
  - Embrace
  - PersonRuns
  - ElevatorNoEntry

- Events in 2010
  - **PeopleMeet**
  - **PeopleSplitUp**
  - **Embrace**
  - PersonRuns

Pair-wise activity

Our Results in TRECVID-ED 2010(1)

- Compared with the best results (according to NDCR) this year

PeopleMeet	#Ref	#Sys	#CorDet	#FA	#Miss	NDCR
PKU-IDM/p-eSur_2	449	156	12	144	437	1.02
PKU-IDM/p-eSur_4	449	4331	11	150	438	1.025
PeopleSplitUp						
PKU-IDM/p-eSur_4	187	167	16	136	171	0.959
PKU-IDM/p-eSur_2	187	157	13	144	174	0.978
Embrace						
IPG-BJTU_5/p-SYS_1	175	64	9	55	166	0.967
PKU-IDM/p-eSur_4	175	925	6	71	169	0.989
PersonRuns						
QMUL-ACTIVA_3	107	360	36	223	71	0.737
PKU-IDM/p-eSur_3	107	2748	2	76	105	1.006

****Systems with 0 correct detection are excluded.***

Our Results in TRECVID-ED 2010(2)

- Compared with our results last year

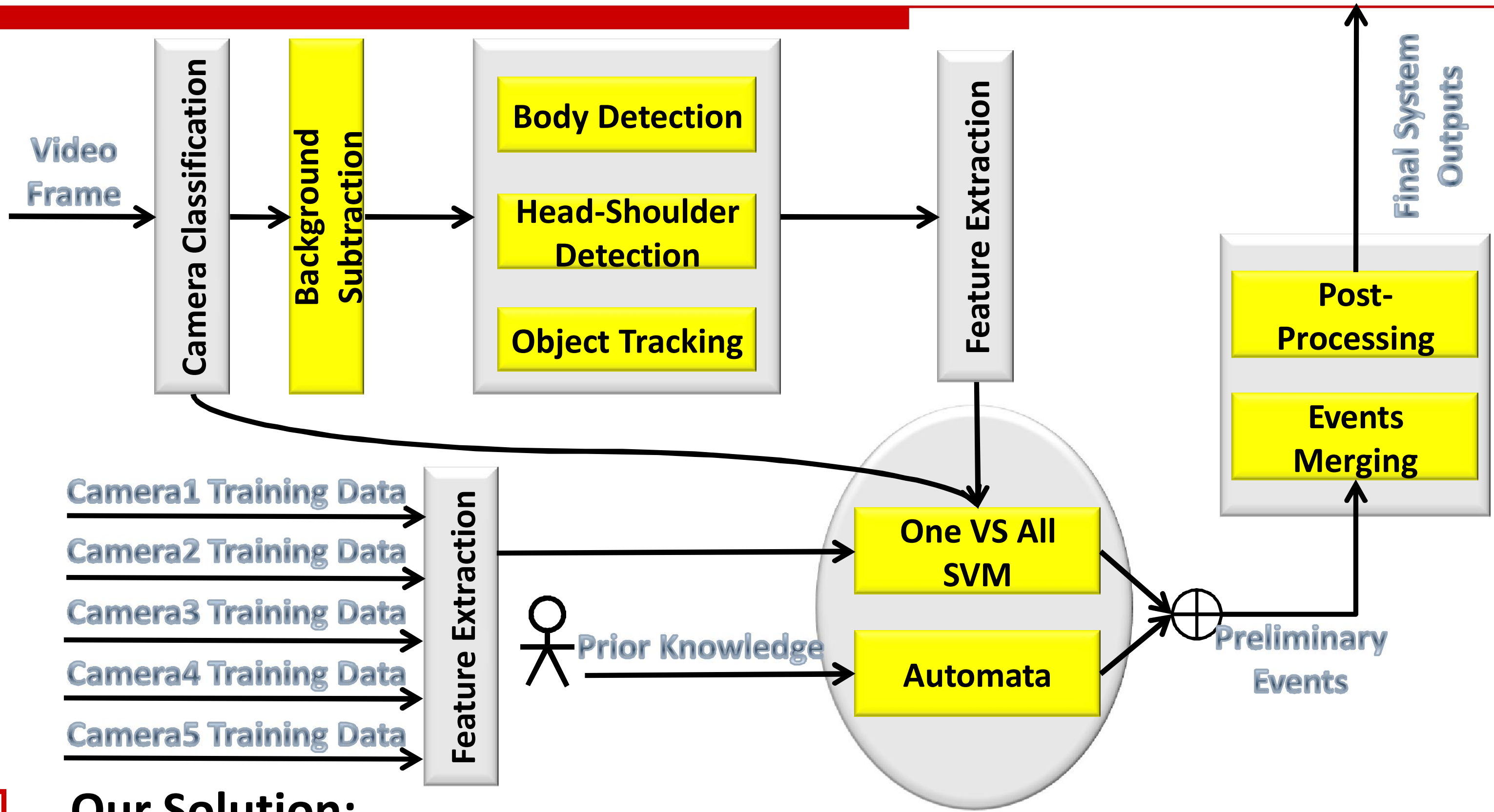
PeopleMeet	#Ref	#Sys	#CorDet	#FA	#Miss	Act.DCR
2009	449	125	<u>**7**</u>	118	442	<u>**1.023**</u>
2010	449	156	<u>**12**</u>	144	437	<u>**1.02**</u>

Improvements on both correct detection rate and Actual DCR!

Embrace						
2009	175	80	<u>**1**</u>	79	174	<u>**1.02**</u>
2010	175	925	<u>**6**</u>	71	169	<u>**0.989**</u>

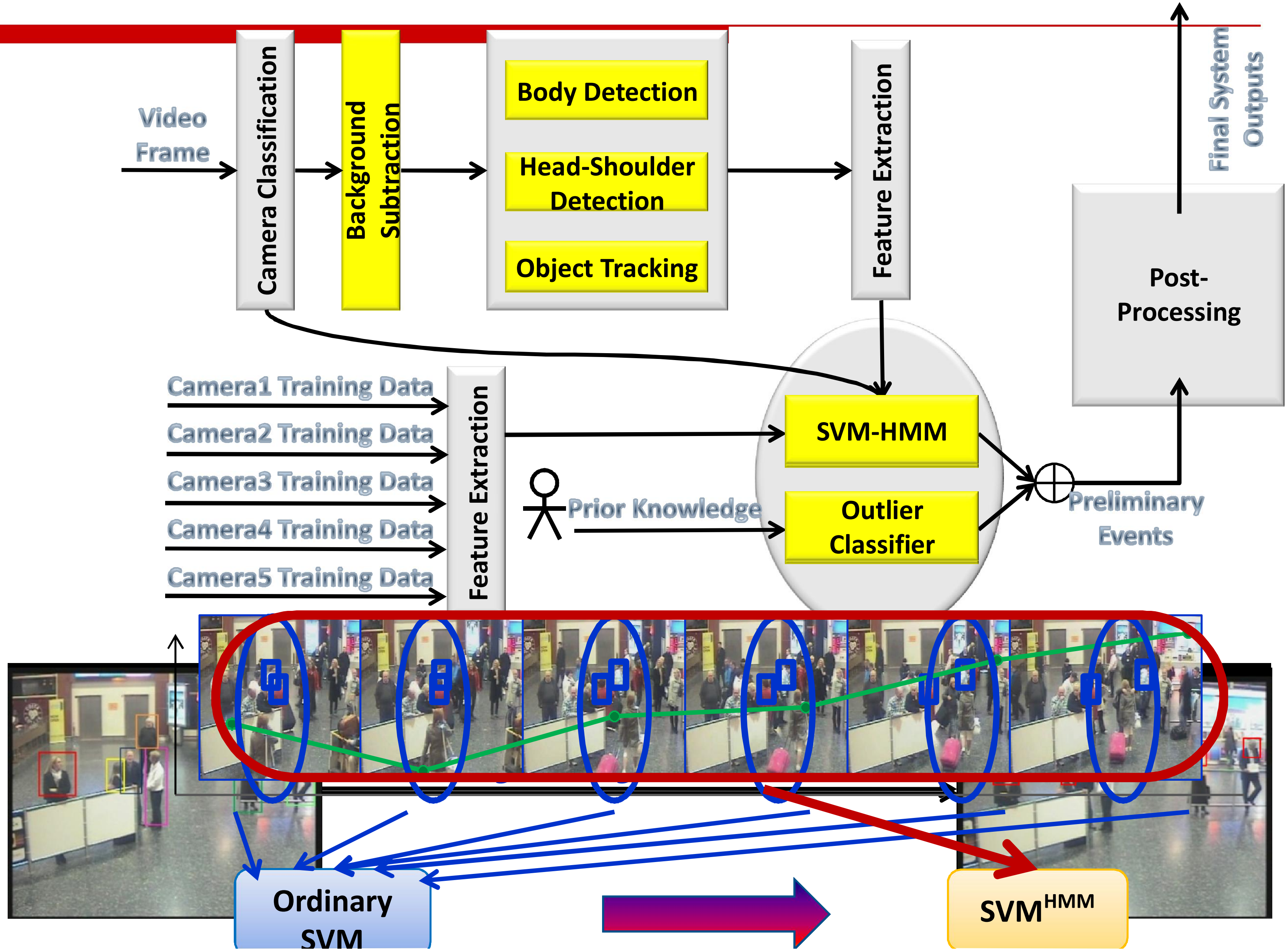
Why?

Our System in 2009: eSur



- **Our Solution:**
  1. Adaptive background modeling
  2. Body and head-shoulder detection and adaboost-based tracking
  3. Ensemble of one-vs.-all SVM and automata-based classifiers
  4. Effective event merging and post-processing

Our System in 2010: eSur v1.2

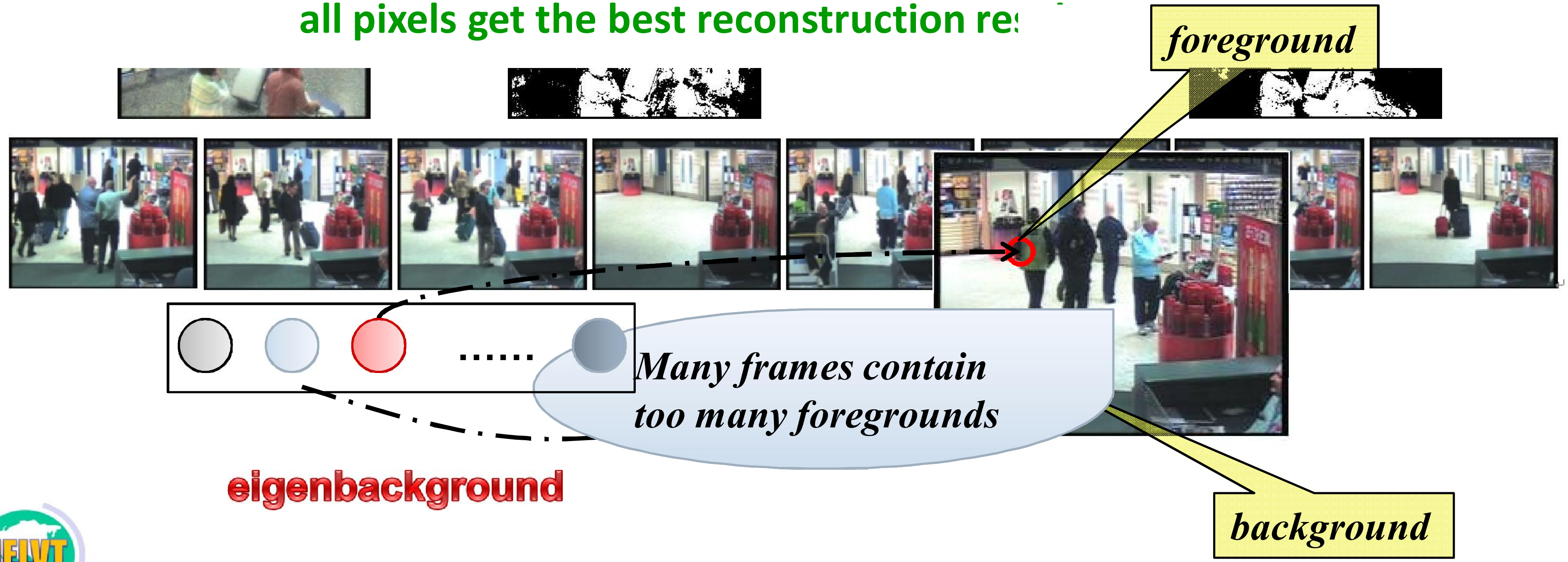


What are the Improvements?

- Background Subtraction
  - Method: *Pixel-level selective eigenbackground*
  - Result: *Better foreground object detection with much lower false alarms in crowded scenes*
- Head-Shoulder Detection
  - Method: *Multi-pose learning for detection*
  - Result : *Greatly boost the recall*
- Event Detection
  - Method: *SVM^{HMM} classifier employed for pair-wise event detection*
  - Result : *More correct detections with less false alarms than last year*

Our Solution (1):
Background Modeling

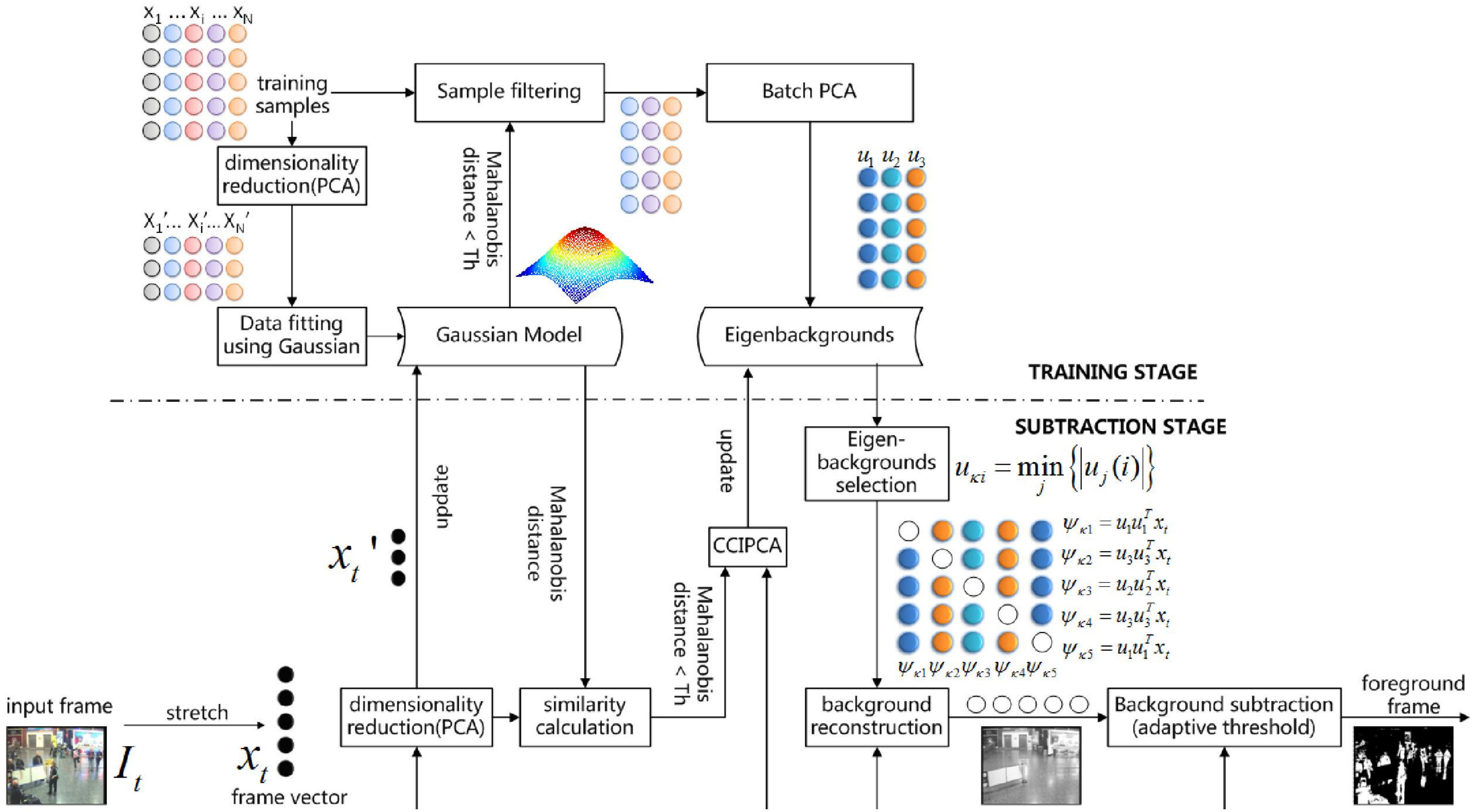
- **Background Modeling in 2009**
  - **Method: Block-wise PCA**
    - **Segment a frame into blocks, and model each block respectively**
  - **Shortcomings**
    - **Background subtraction is performed on frame level. As such, not all pixels get the best reconstruction re**



Selective Eigenbackgounds (1)

Main Idea

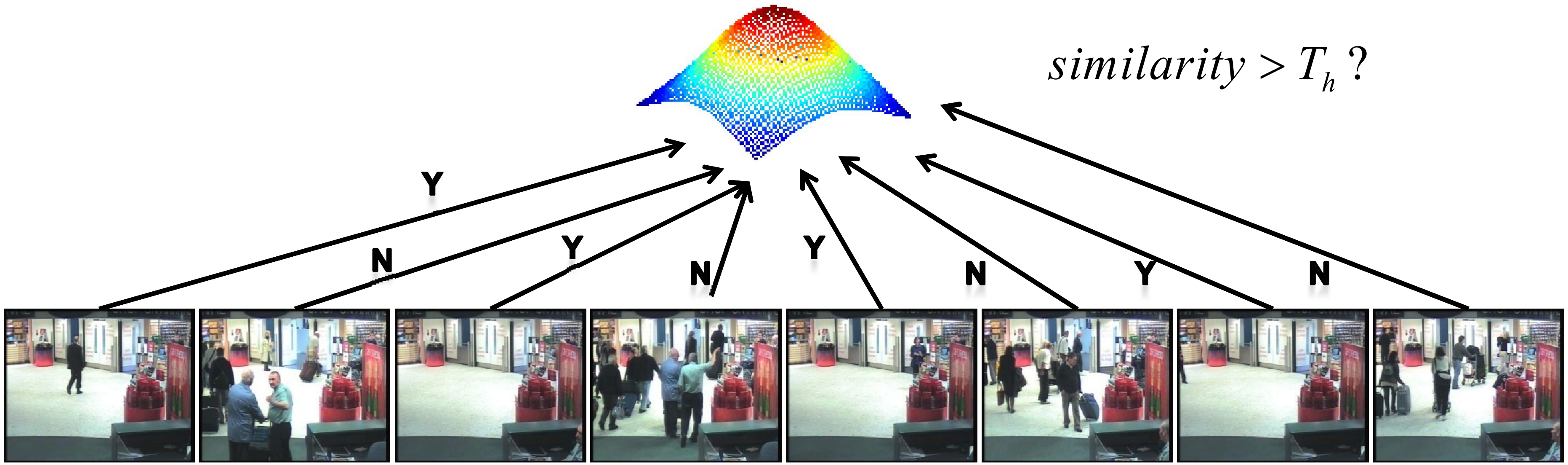
- Select frames with fewer foregrounds to train eigenbackgrounds
- Background reconstruction is performed selectively on pixel level
- Adaptive thresholding strategy is employed for background subtraction



Selective Eigenbackgounds (2)

Frame Selection for Background Modeling

- A Gaussian model is used to describe the crowd density of a scene
- Select frames with fewer foregrounds for background initialization and update by judging the similarity between frames and GMM

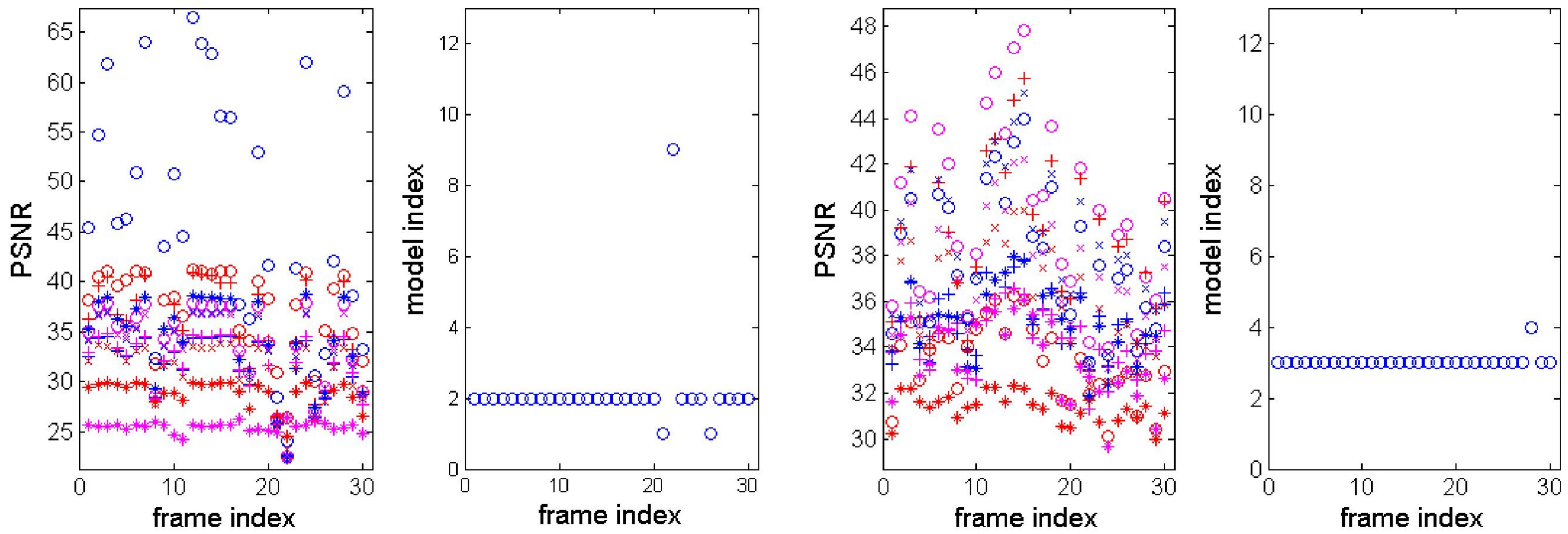


High-similar frames selected for background initialization and update

Selective Eigenbackgounds (3)

PSNR-based Model Selection

- Multiple background models are trained
- Model Selection is used to choose the background model in the database that most fits the observed scene.
  - Peak signal-to-noise ratio

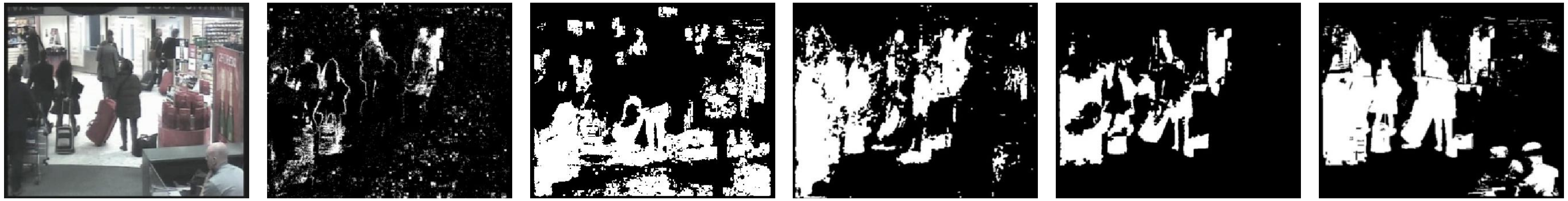


(a) experiment on MCTTR0201c (b) experiment on MCTTR0305d

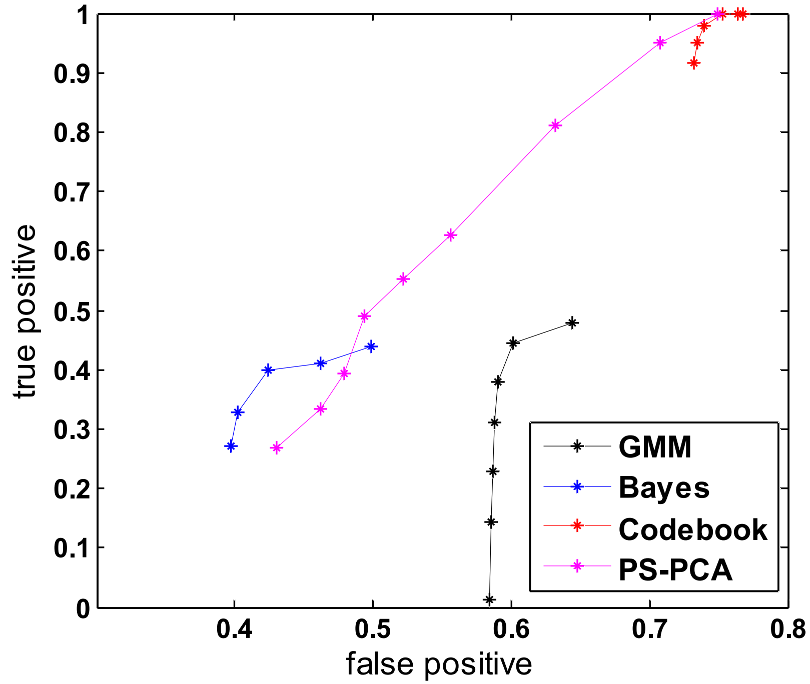
model selection experiment: For each frame, the PSNRs between itself and the reconstructed background images using the trained background models are computed. Then a model can be selected according to the maximum PSNR. Finally, the most suitable model can be determined by voting on the selection results from the 30 frames.

Experimental Results (1)

- Compared with several state-of-the-art methods

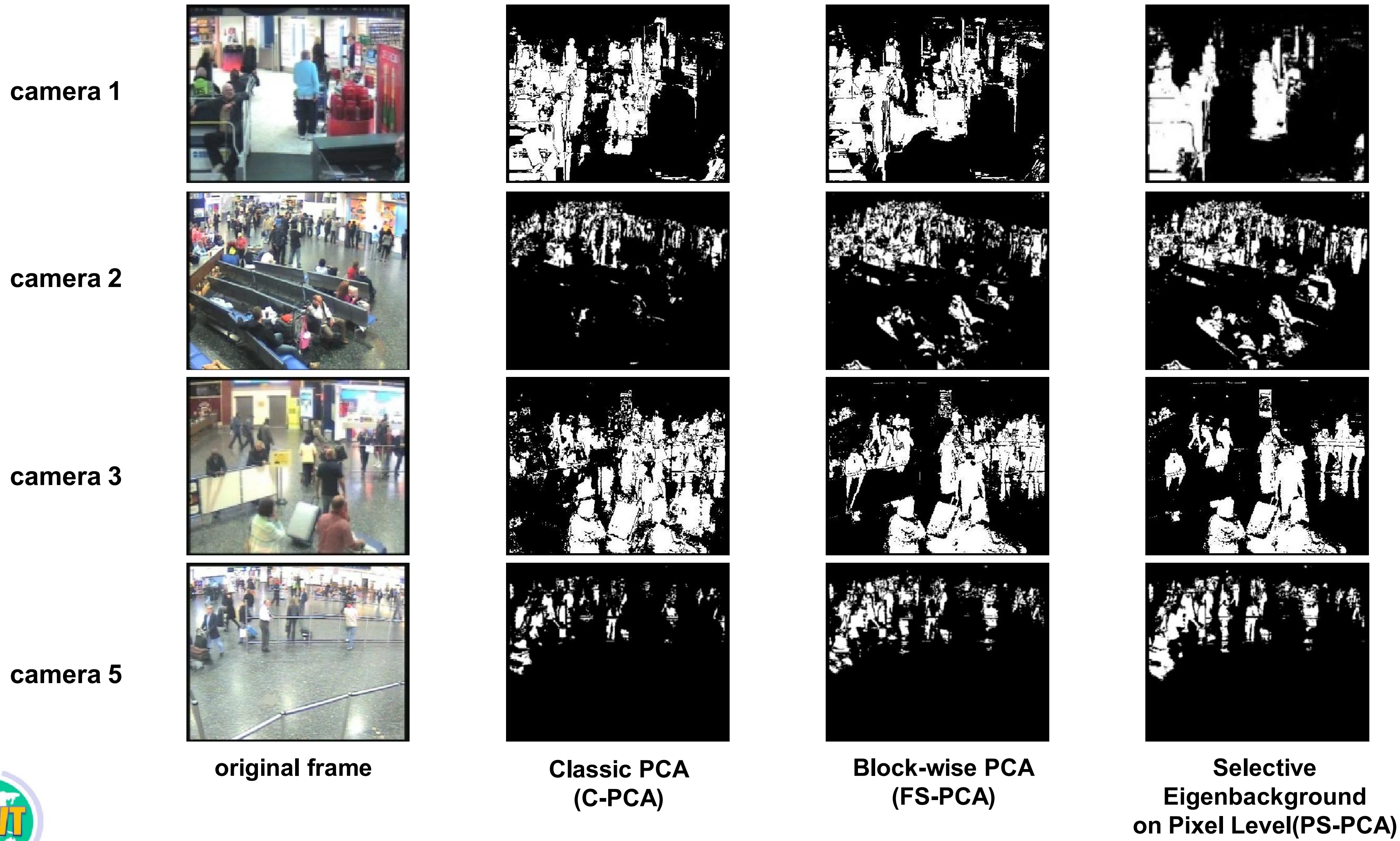


original frame | GMM [Stauffer,1999] | KDE [Elgammal,2000] | Codebook [Kim, 2005] | Bayes method [Li, 2003] | Our method



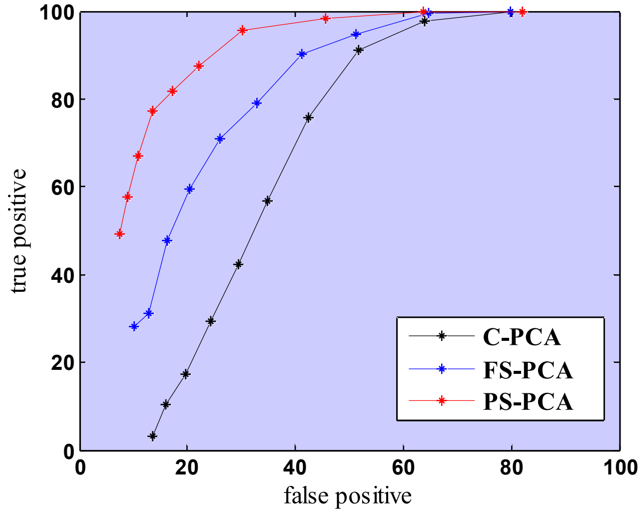
Experimental Results (2)

- Compared with other eigenbackground methods

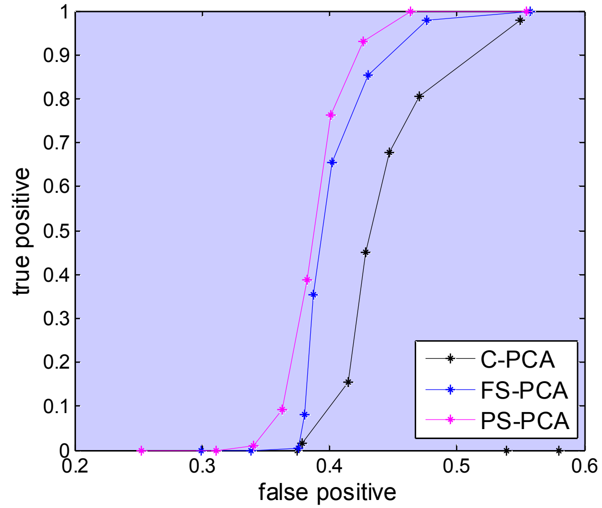


Experimental Results (3)

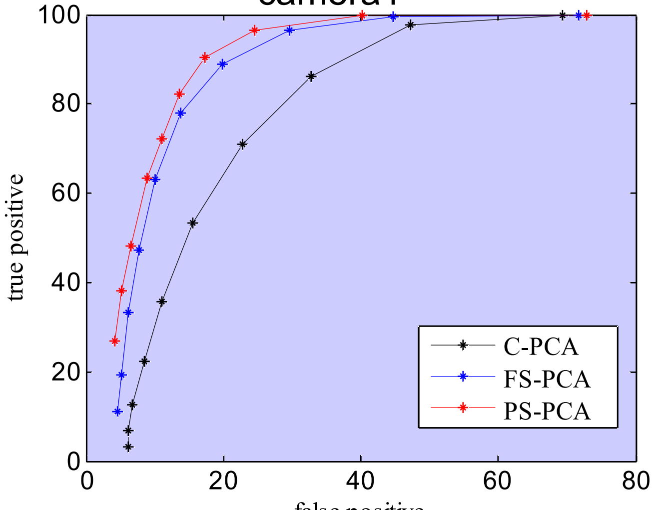
- Compared with other eigenbackground methods



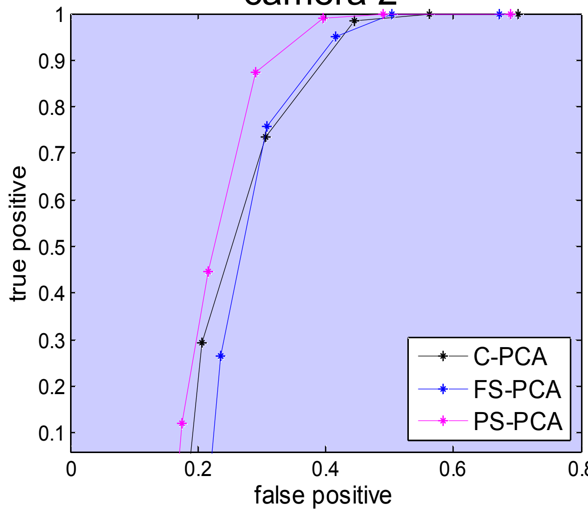
camera1



camera 2



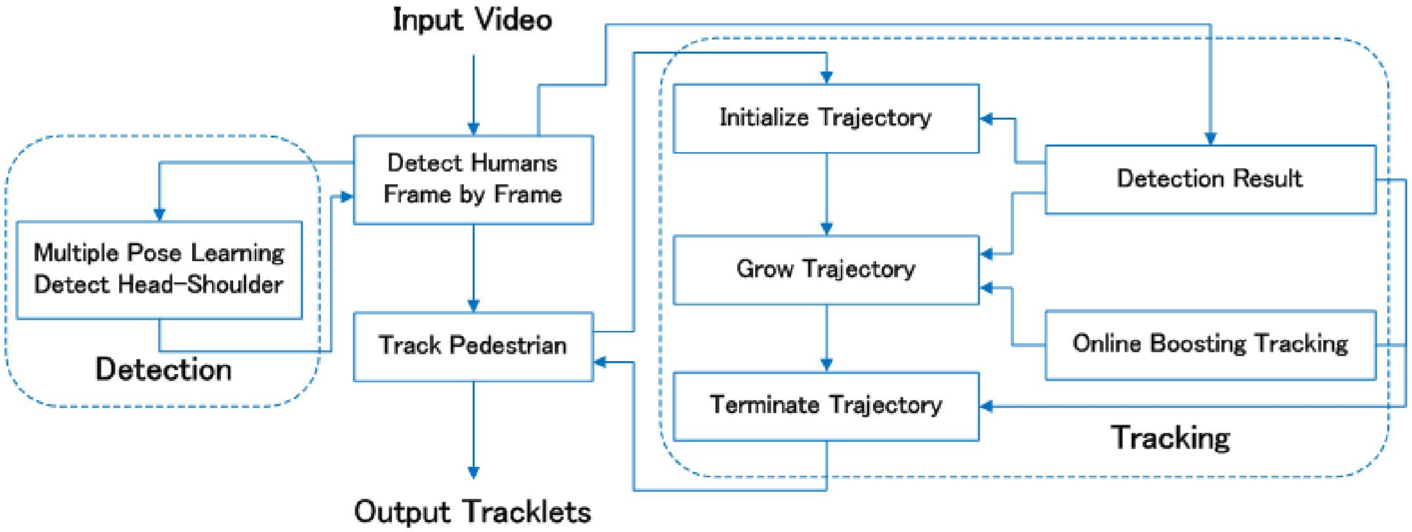
camera 3



camera 5

Our Solution (2): MPL Detection and Tracking

- Head-shoulder Detection:
  - Feature: Histogram of oriented gradients (HOG)
  - Classifier: Multiple pose learning [1]
- Tracking
  - Online boosting [2]
  - Combining color similarity to reduce drift



[1] Boris Babenko, Piotr dollar et al, Simultaneous Learning and Alignment: Multi-Instance and Multi-Pose Learning, ECCV, 2008.

[2] Helmut Grabner et al, Online Boosting and Vision, CVPR,2006.

Multiple Pose Learning

- The detector works best when trained with images that come from *a single coherent group* and *lie in approximate correspondence* [1].
- Issue: Data Confusion

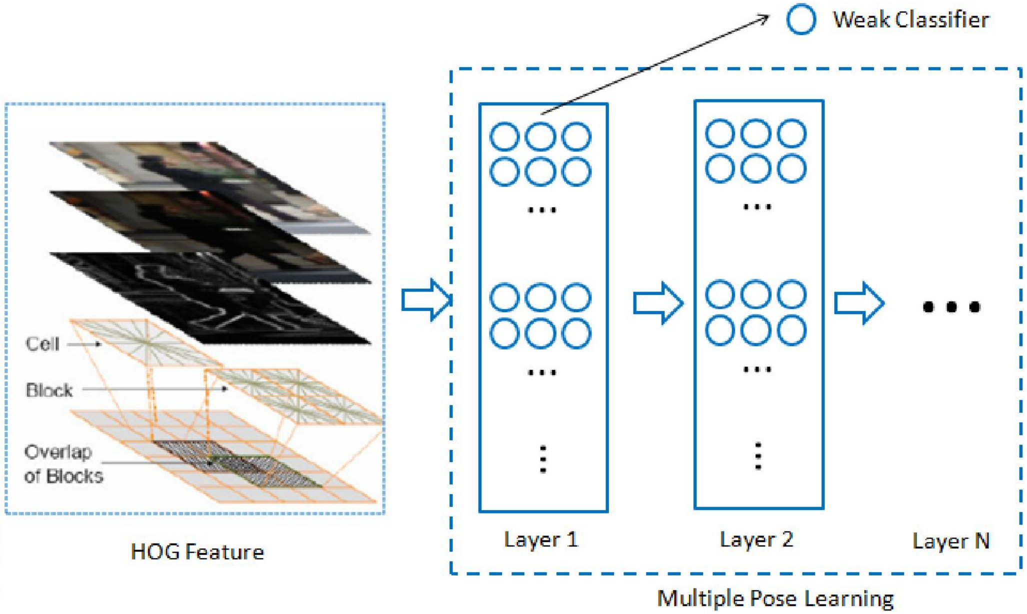
Intra-class diversification
vs.
Inter-class correlation

- Solution: Data Alignment
  - To split data into groups and train classifiers for each

[1] Boris Babenko, Piotr dollar et al, Simultaneous Learning and Alignment: Multi-Instance and Multi-Pose Learning, ECCV, 2008.

Cascaded Classifiers of MPL

- Detection Framework
  - Multiple Pose Learning : Simultaneously group the positive data, and train classifiers for each of the K groups by combining weak classifiers
    - Each positive sample is scored by K weak classifiers from different aspects
  - Cascaded Classifiers
    - Classifiers are combined using a boosting manner



FOR t= 1 TO DO

FOR k=1 TO K DO

Compute weights $\omega_i^k = -\frac{\partial \varsigma}{\partial H^k(x_i)}$

Train the best weak classifier with the current weights

$$h_t^k = \arg\min_h \sum_i 1(h(x_i) \neq y_i)\left|\omega_i^k\right|$$

Find α_t^k via line search to minimize cost

$$\alpha_t^k = \arg\min_\alpha \varsigma(\ldots, H^k(x) + \alpha h_t^k, \ldots)$$

Update strong classifier

$$H^k(x) \leftarrow H^k(x) + \alpha_t^k h_t^k$$

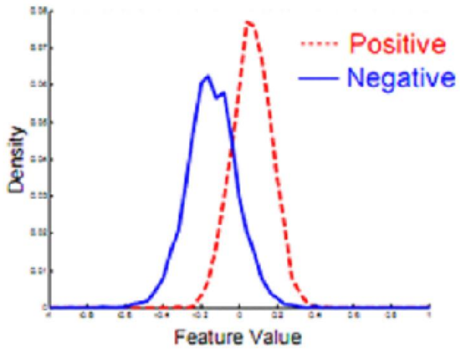
END FOR

END FOR

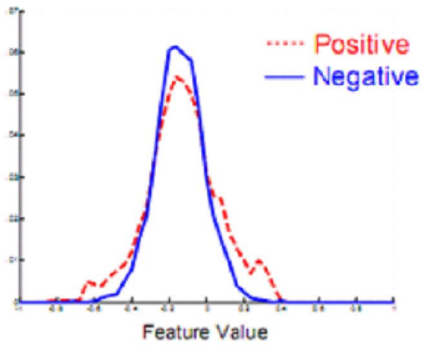
Define probability as a softmax of probabilities determined by each classifier and optimize the loss function (i.e., the negative log likelihood), where derivative of the loss function gives the instance weights for each classifier

Weak Classifier

Piecewise Function

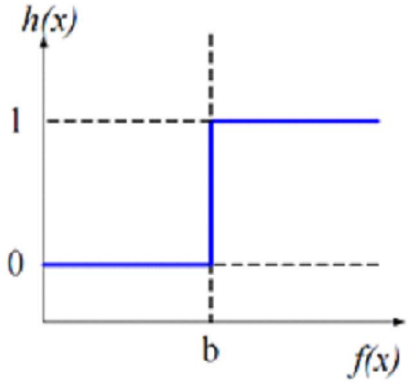


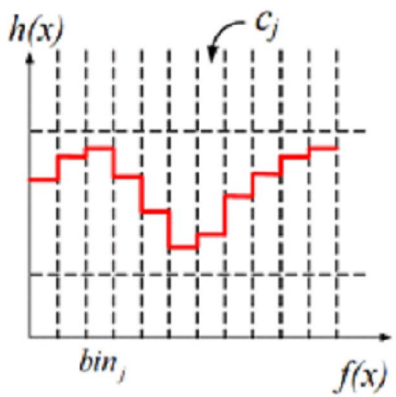
Linear separable case



Linear non-separable case

Sample of feature distribution

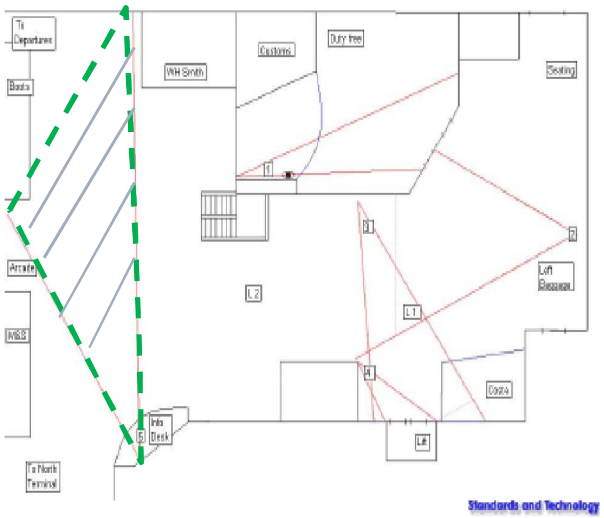




Decision tree and piecewise function

Cascaded Classifiers of MPL

- Adjust the detector searching scales



Experimental Results

- On a labeled TRECVID 2008 corpus

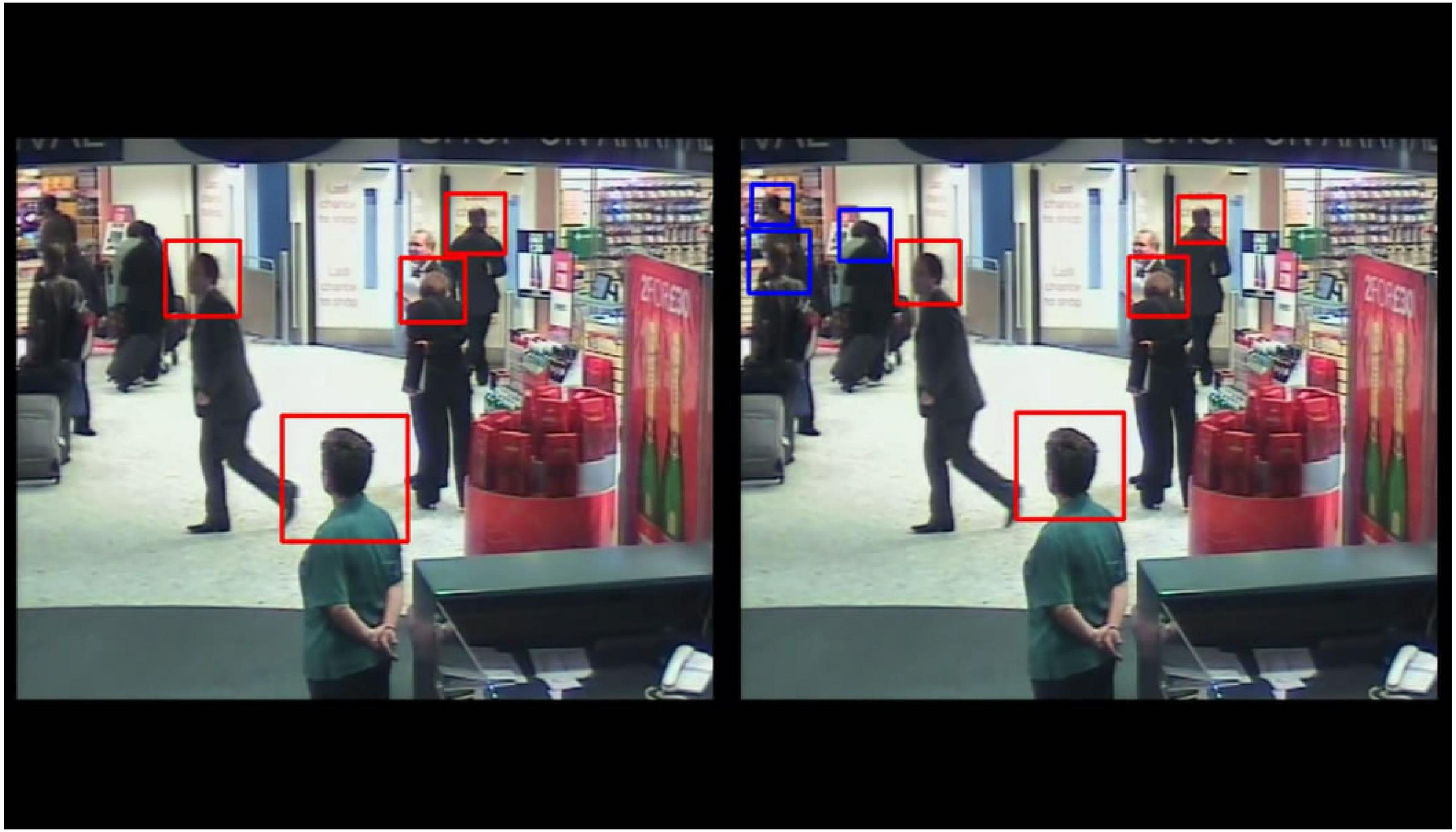
Camera1	Recall	Precision	F
Cascade HOG	33.5%	88.8%	0.4734
MPL	53.9%	79.6%	0.6429

Camera2	Recall	Precision	F
Cascade HOG	24.3%	81.6%	0.3745
MPL	56.0%	77.3%	0.6495

Camera3	Recall	Precision	F
Cascade HOG	30.5%	72.8%	0.4299
MPL	42.9%	66.7%	0.5222

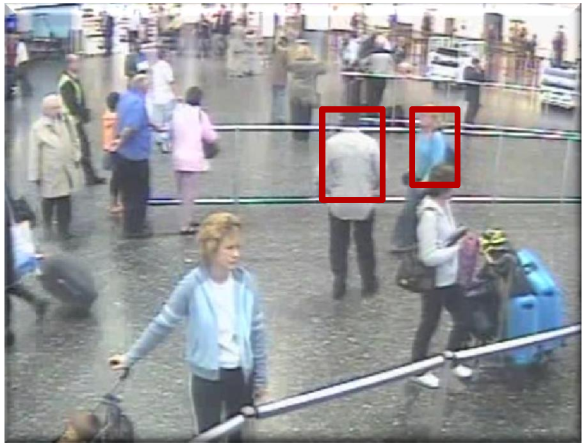
Camera5	Recall	Precision	F
Cascade HOG	38.5%	66.2%	0.4869
MPL	46.8%	75.7%	0.5783

Visualized Explanation

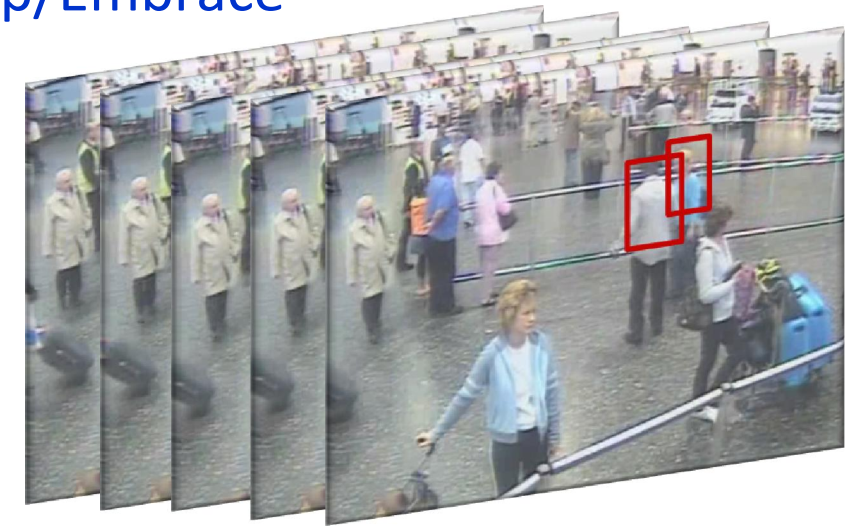


Our Solution (3): Sequential Learning for Event Detection

- Event Analysis based on Sequential Learning
  - Video events are inherently time sequential patterns
  - Model the activity as sequence structure and consider the information in and between frames
  - Our current work focuses on pair activities, e.g. PeopleMeet/SplitUp/Embrace

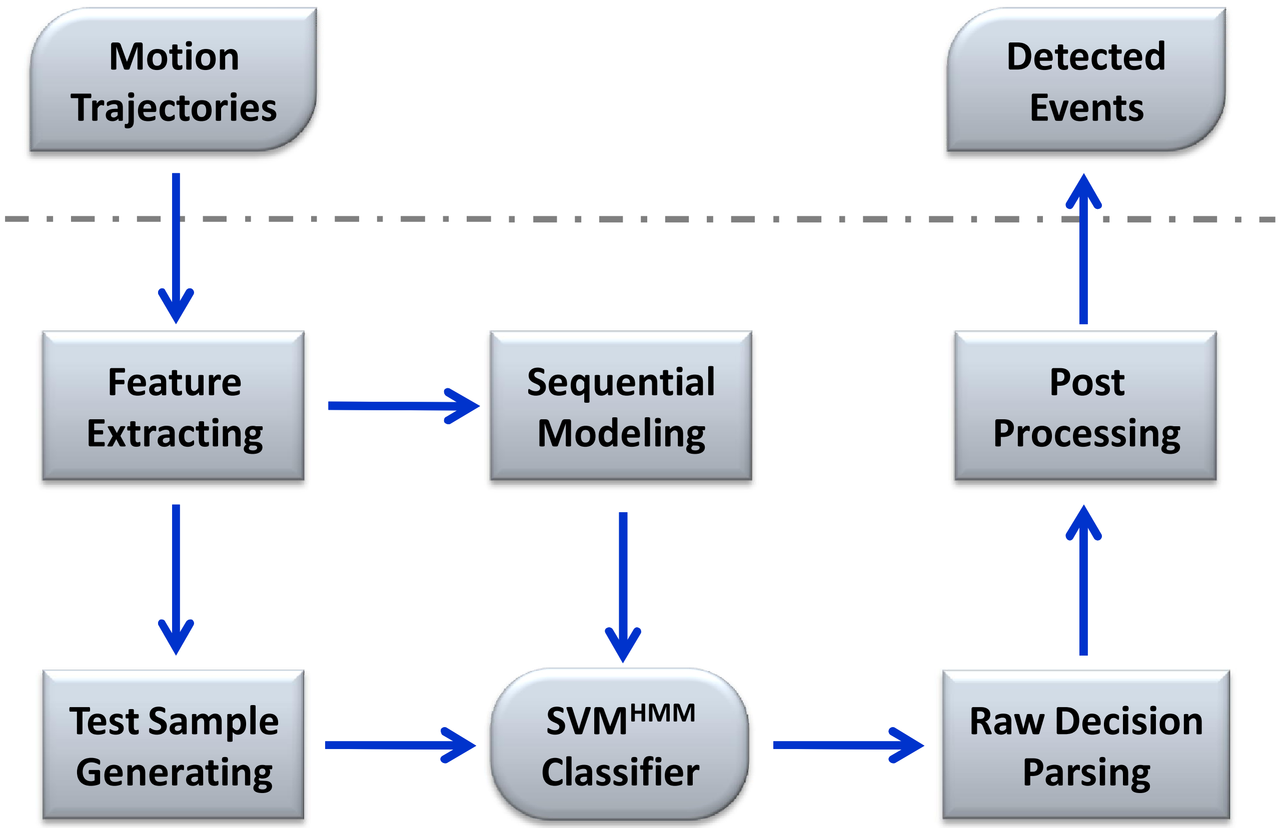


Meet, SplitUp or just Stand&Talk?



PeopleMeet !

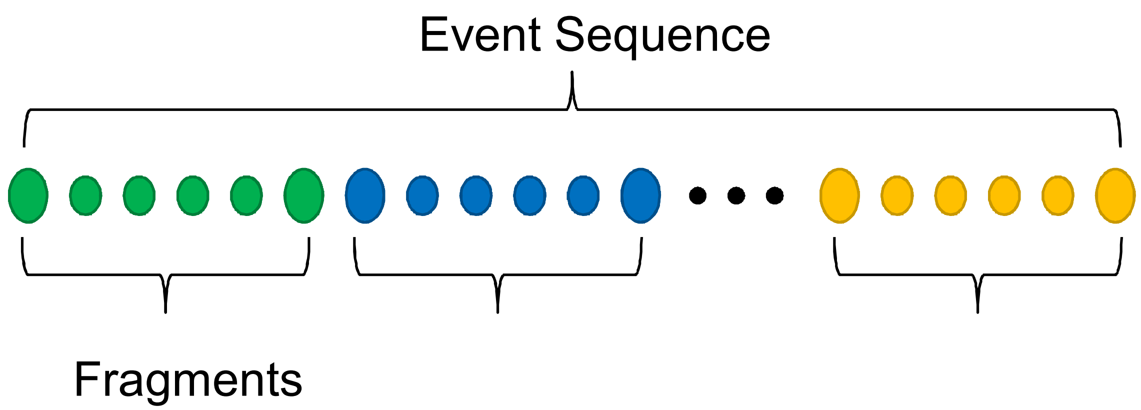
Detection Framework



▲ In our implemented system,
classifier is trained for each type of event

Sequential Learning for Event Detection (1)

- Structural Modeling
  - Treat event video clips as holistic frame sequences
  - A small number of adjacent frames make up a fragment
  - Model the event sequence as a set of contiguous fragments



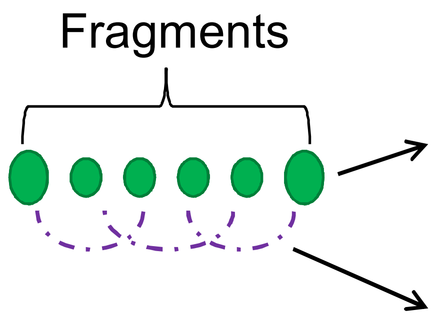
Sequential Learning for Event Detection (2)

- Features of Fragments
  - Describe frames of fragment and represent the fragment
  - Trajectory based motion and pair features:
    - Absolute velocity, acceleration
    - Angular separation of moving directions
    - Distance between pair of persons
    - Statistics of the features within several adjacent frames
      - The mean, variation, trend of distances between persons

$$Ang(P_1,P_2) = \left\{ i \,\middle|\, \left|\theta_{P_1} - \theta_{P_2}\right| \in \left[(i-1) * \frac{\pi}{4}, i * \frac{\pi}{4}\right), i \in Z^+ \right\}$$

$$Dist(P_1,P_2) = \left\| pos_{P_1} - pos_{P_2} \right\|$$

$$T = \frac{1}{N} \sum_{i=1}^{N-1} \frac{1}{\overline{Dist}} (Dist_i - Dist_{i+1})$$



Features extracted from frames describe the basic information of event

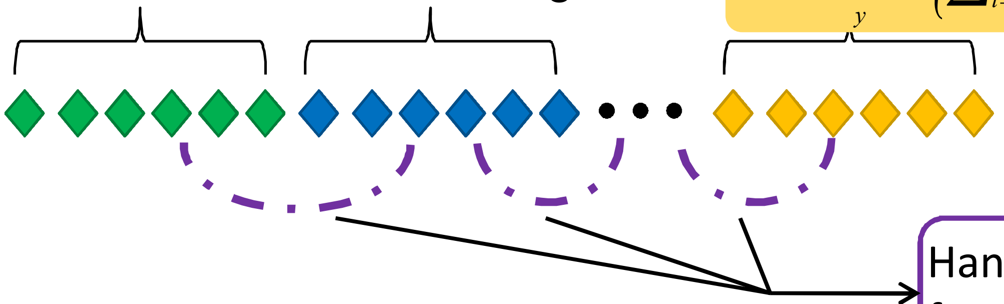
Statistics employs correlation within fragment

Sequential Learning for Event Detection (3)

- Sequence Learning
  - Represent events as feature sequences, but not concatenated feature vectors
  - Dynamics of the pattern within an event is modeled by Hidden Markov Model[1]
  - Learning and classification processes are performed by an implementation of structural SVM, SVM^HMM[2]

Features of Fragments

$$y = \arg\max_{y} \left\{ \sum_{i=1..l} \left[\sum_{j=1..k} (x_i \bullet w_{y_{i-j} \dots y_i}) + \phi_{trans}(y_{i-j,\dots,y_i}) \bullet w_{trans} \right] \right\}$$



Handling dependencies between adjacent fragments using Viterbi decoding

[1] Yasemin Altun, Ioannis Tsochantaridis and Thomas Hofmann. Hidden Markov Support Vector Machines. International Conference on Machine Learning (ICML), 2003.
[2] Thorsten Joachims, Sequence Tagging with Structural Support Vector Machines, http://www.cs.cornell.edu/People/tj/svm_light/svm_hmm.html

Sequential Learning for Event Detection (4)

- Decision making and Post Processing
  - Divide videos for detection into test samples using sliding window strategy
  - Sequential results are generated by SVM^{HMM} classifiers
  - Transform classification sequence to raw decision with voting
  - Exploit priors for post-processing

▲ numbers stand for event class labels

Experimental Results

- Evaluation on 10 hours data from TREVID-ED 2008 corpus

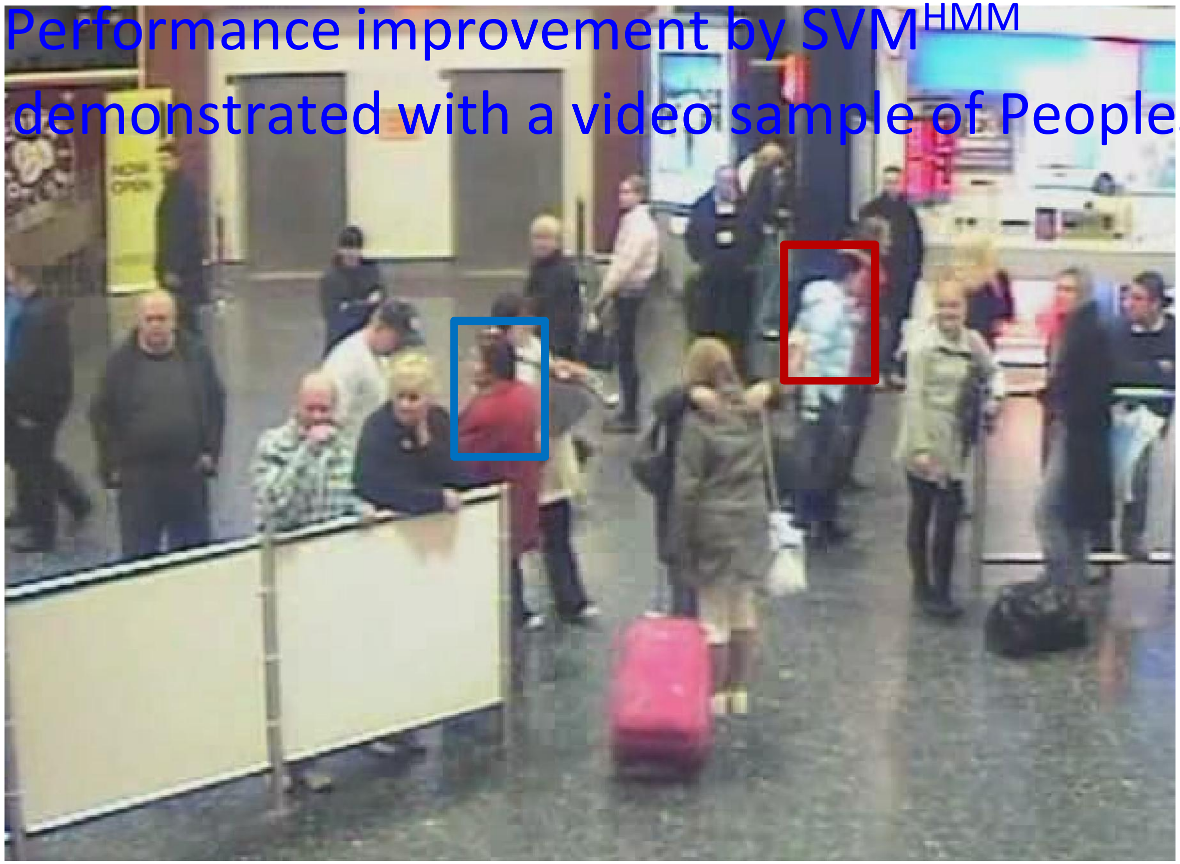
event	#Ref		#Sys	#CorDet	#FA	#Miss	DCR	NDCR
PeopleMeet	298	★	54	7	47	291	198.21	1.000
		◇	29	2	27	296	200.34	1.007
PeopleSplitUp	152	★	81	7	74	145	195.23	0.991
		◇	21	0	21	152	201.31	1.011
Embrace	116	★	82	5	77	111	196.19	0.995
		◇	44	1	43	115	200.96	1.000

★ is results of sequential learning, SVM^{HMM}
◇ is results of last year's ordinary SVM

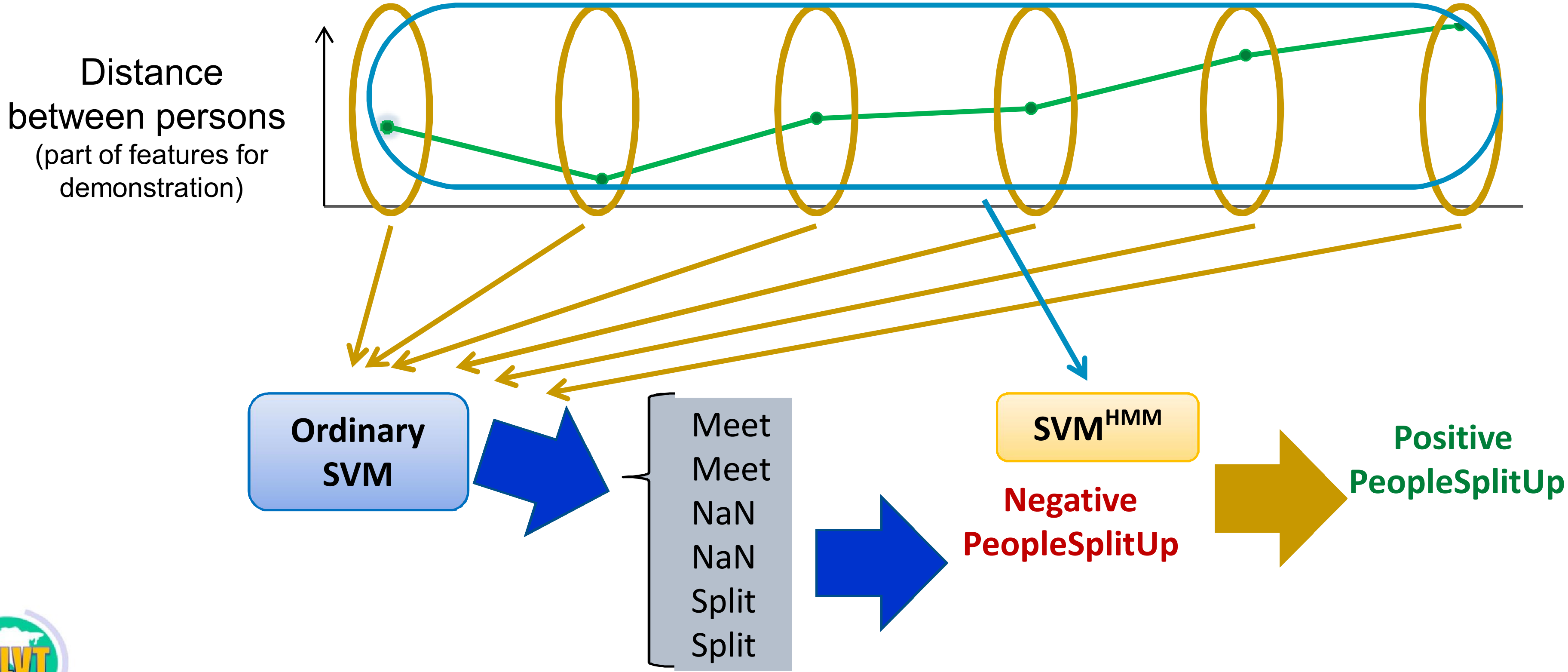
Obtain performance improvement, especially on the number of correct detection

Visualized Explanation

- Experiments
  - Performance improvement by SVM^{HMM} demonstrated with a video sample of PeopleSplitUp

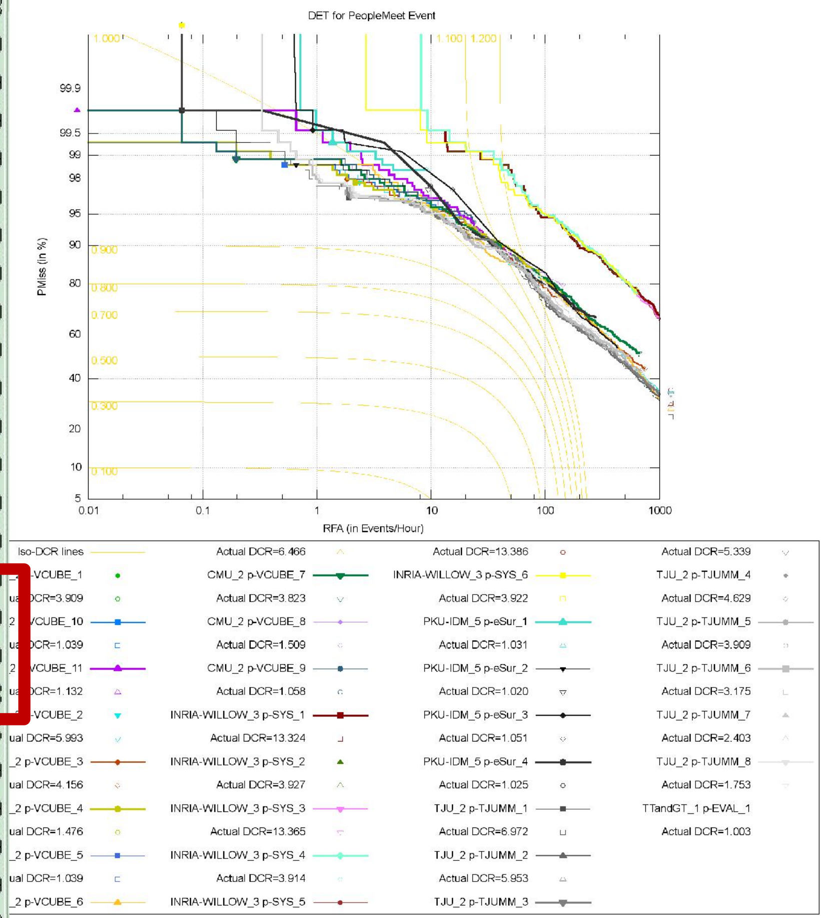


Visualized Explanation



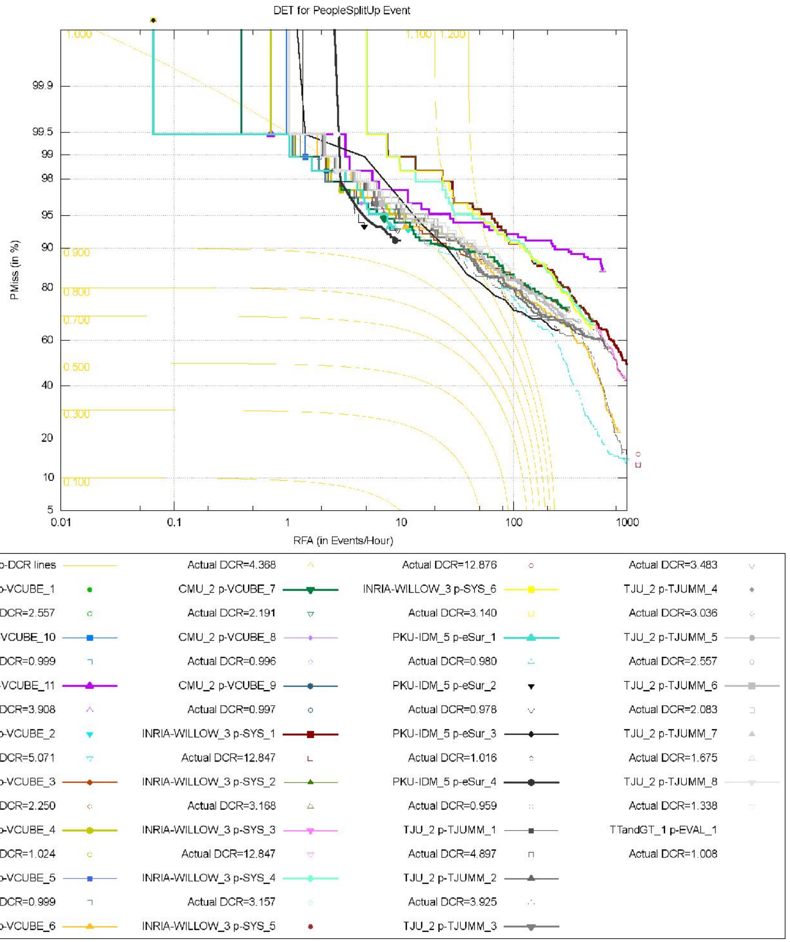
Evaluation Results – PeopleMeet

Analysis Report	#Ref	#Sys	#CorDet	#FA	#Miss	Act. RFA	Act. PMiss	Act. DCR	Min RFA	Min PMiss	Min DCR
CMU_2 / p-VCUBE_1	449	10841	253	10588	196	694.422	0.436	3.909	2.099	0.969	0.979
CMU_2 / p-VCUBE_10	449	305	24	281	425	18.430	0.947	1.039	0.525	0.987	0.989
CMU_2 / p-VCUBE_11	449	854	58	796	391	52.206	0.871	1.132	0.000	0.998	0.998
CMU_2 / p-VCUBE_2	449	17547	298	17249	151	1131.288	0.336	5.993	2.361	0.978	0.990
CMU_2 / p-VCUBE_3	449	11563	249	11314	200	742.037	0.445	4.156	1.836	0.980	0.989
CMU_2 / p-VCUBE_4	449	2261	104	2157	345	141.468	0.768	1.476	2.164	0.978	0.989
CMU_2 / p-VCUBE_5	449	305	24	281	425	18.430	0.947	1.039	0.525	0.987	0.989
CMU_2 / p-VCUBE_6	449	19215	327	18888	122	1238.783	0.272	6.466	0.197	0.989	0.990
CMU_2 / p-VCUBE_7	449	10307	218	10089	231	661.694	0.514	3.823	0.197	0.989	0.990
CMU_2 / p-VCUBE_8	449	2260	91	2169	358	142.255	0.797	1.509	0.197	0.989	0.990
CMU_2 / p-VCUBE_9	449	388	27	361	422	23.676	0.940	1.058	0.197	0.989	0.990
INRIA-WILLOW_3 / p-SYS_1	449	40045	316	39729	133	2605.655	0.296	13.325	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_2	449	9696	99	9597	350	629.426	0.779	3.927	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_3	449	40045	300	39745	149	2606.704	0.332	13.365	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_4	449	9696	104	9592	345	629.098	0.768	3.914	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_5	449	40045	292	39753	157	2607.229	0.350	13.386	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_6	449	9696	101	9595	348	629.295	0.775	3.921	0.066	1.000	1.000
PKU-IDM_5 / p-eSur_1	449	148	7	141	442	9.248	0.984	1.031	1.377	0.993	1.000
PKU-IDM_5 / p-eSur_2	449	156	12	144	437	9.444	0.973	1.020	0.656	0.987	0.990
PKU-IDM_5 / p-eSur_3	449	6781	12	236	437	15.478	0.973	1.051	0.918	0.996	1.000
PKU-IDM_5 / p-eSur_4	449	4331	11	150	438	9.838	0.976	1.025	0.066	0.998	0.998
TJU_2 / p-TJUMM_1	449	20859	340	20519	109	1345.753	0.243	6.971	1.902	0.967	0.976
TJU_2 / p-TJUMM_2	449	17596	320	17276	129	1133.059	0.287	5.953	1.968	0.969	0.979
TJU_2 / p-TJUMM_3	449	15568	300	15268	149	1001.363	0.332	5.339	1.968	0.969	0.979
TJU_2 / p-TJUMM_4	449	13278	284	12994	165	852.221	0.367	4.629	2.033	0.969	0.979
TJU_2 / p-TJUMM_5	449	10841	253	10588	196	694.422	0.436	3.909	2.099	0.969	0.979
TJU_2 / p-TJUMM_6	449	8378	224	8154	225	534.786	0.501	3.175	2.099	0.969	0.979
TJU_2 / p-TJUMM_7	449	5814	197	5617	252	368.395	0.561	2.403	2.164	0.969	0.980
TJU_2 / p-TJUMM_8	449	3482	152	3330	297	218.400	0.661	1.753	2.230	0.969	0.980
TTandGT_1 / p-EVAL_1	449	8	0	8	449	0.525	1.000	1.003	0.525	1.000	1.003



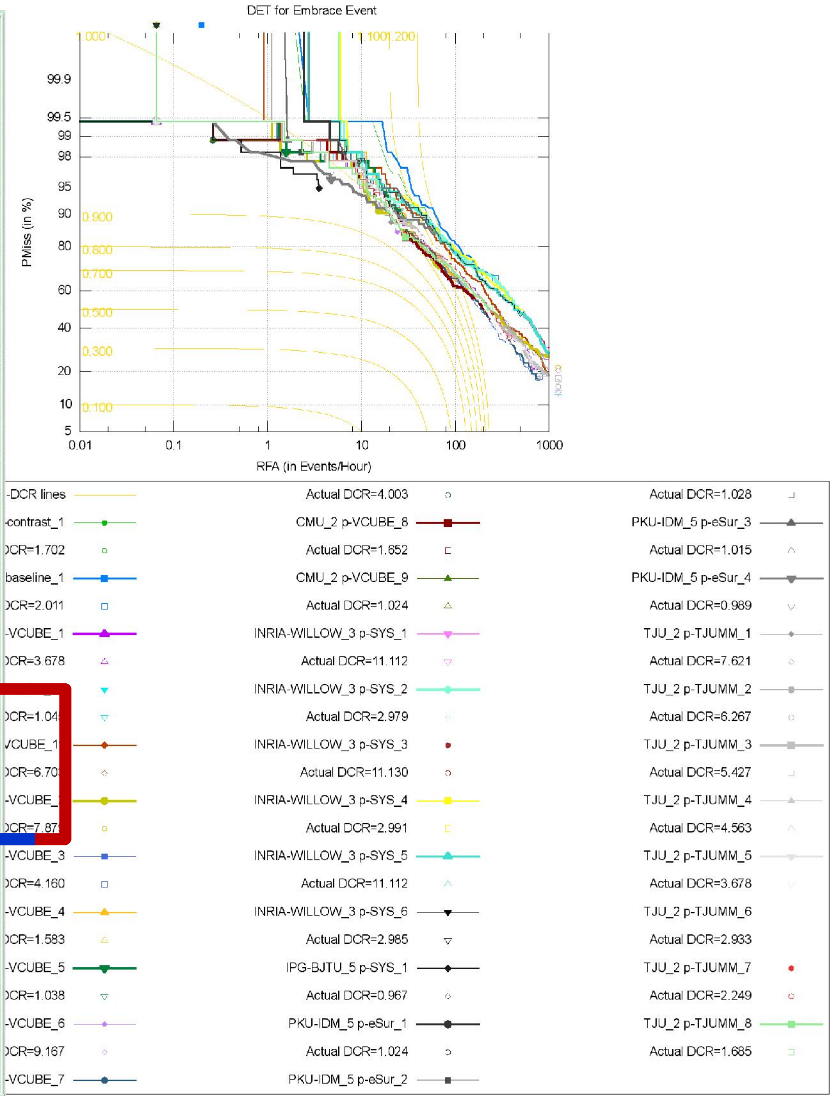
Evaluation Results – PeopleSplitUp

Analysis Report	#Ref	#Sys	#CorDet	#FA	#Miss	Act. RFA	Act. PMiss	Act. DCR	Min RFA	Min PMiss	Min DCR
CMU_2 / p-VCUBE_1	187	5787	60	5727	127	375.609	0.679	2.557	2.689	0.984	0.997
CMU_2 / p-VCUBE_10	187	31	2	29	185	1.902	0.989	0.999	1.443	0.989	0.997
CMU_2 / p-VCUBE_11	187	9351	28	9323	159	611.456	0.850	3.907	0.721	0.995	0.998
CMU_2 / p-VCUBE_2	187	15201	161	15040	26	986.409	0.139	5.071	11.674	0.930	0.989
CMU_2 / p-VCUBE_3	187	4713	52	4661	135	305.695	0.722	2.250	7.149	0.952	0.988
CMU_2 / p-VCUBE_4	187	265	11	254	176	16.659	0.941	1.024	3.017	0.973	0.988
CMU_2 / p-VCUBE_5	187	31	2	29	185	1.902	0.989	0.999	1.443	0.989	0.997
CMU_2 / p-VCUBE_6	187	12779	145	12634	42	828.610	0.225	4.368	11.150	0.936	0.992
CMU_2 / p-VCUBE_7	187	4514	51	4463	136	292.709	0.727	2.191	7.214	0.947	0.983
CMU_2 / p-VCUBE_8	187	281	17	264	170	17.315	0.909	0.996	4.525	0.963	0.985
CMU_2 / p-VCUBE_9	187	42	3	39	184	2.558	0.984	0.997	2.230	0.984	0.995
INRIA-WILLOW_3 / p-SYS_1	187	38949	163	38786	24	2543.808	0.128	12.847	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_2	187	7650	60	7590	127	497.796	0.679	3.168	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_3	187	38949	163	38786	24	2543.808	0.128	12.847	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_4	187	7650	62	7588	125	497.664	0.668	3.157	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_5	187	38949	158	38791	29	2544.135	0.155	12.876	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_6	187	7650	65	7585	122	497.468	0.652	3.140	0.066	1.000	1.000
PKU-IDM_5 / p-eSur_1	187	147	12	135	175	8.854	0.936	0.980	8.067	0.936	0.976
PKU-IDM_5 / p-eSur_2	187	157	13	144	174	9.444	0.930	0.978	4.788	0.936	0.960
PKU-IDM_5 / p-eSur_3	187	3848	11	228	176	14.954	0.941	1.016	0.066	1.000	1.000
PKU-IDM_5 / p-eSur_4	187	167	16	136	171	8.920	0.914	0.959	8.920	0.914	0.959
TJU_2 / p-TJUMM_1	187	14601	157	14444	30	947.320	0.160	4.897	5.771	0.963	0.991
TJU_2 / p-TJUMM_2	187	10303	80	10223	107	670.483	0.572	3.925	6.034	0.963	0.993
TJU_2 / p-TJUMM_3	187	8854	74	8780	113	575.843	0.604	3.483	6.165	0.963	0.993
TJU_2 / p-TJUMM_4	187	7421	70	7351	117	482.121	0.626	3.036	2.689	0.984	0.997
TJU_2 / p-TJUMM_5	187	5787	60	5727	127	375.609	0.679	2.557	2.689	0.984	0.997
TJU_2 / p-TJUMM_6	187	4290	57	4233	130	277.624	0.695	2.083	6.886	0.963	0.997
TJU_2 / p-TJUMM_7	187	2784	42	2742	145	179.836	0.775	1.675	2.755	0.984	0.998
TJU_2 / p-TJUMM_8	187	1515	28	1487	159	97.526	0.850	1.338	2.755	0.984	0.998
TTandGT_1 / p-EVAL_1	187	43	1	42	186	2.755	0.995	1.008	2.755	0.995	1.008



Evaluation Results - Embrace

Analysis Report	#Ref	#Sys	#CorDet	#FA	#Miss	Act. RFA	Act. PMiss	Act. DCR	Min RFA	Min PMiss	Min DCR
BUPT-MCPRL_2010092204 / c-contrast_1	175	3155	55	3100	120	203.316	0.686	1.702	0.197	1.000	1.001
BUPT-MCPRL_2010092204 / p-baseline_1	175	4171	59	4112	116	269.688	0.663	2.011	0.197	1.000	1.001
CMU_2 / p-VCUBE_1	175	10691	137	10554	38	692.192	0.217	3.678	0.066	0.994	0.995
CMU_2 / p-VCUBE_10	175	525	21	504	154	33.055	0.880	1.045	1.574	0.983	0.991
CMU_2 / p-VCUBE_11	175	20080	146	19934	29	1307.386	0.166	6.703	1.377	0.989	0.996
CMU_2 / p-VCUBE_2	175	23500	137	23363	38	1532.279	0.217	7.878	15.347	0.909	0.985
CMU_2 / p-VCUBE_3	175	12270	143	12127	32	795.358	0.183	4.160	1.508	0.983	0.990
CMU_2 / p-VCUBE_4	175	3454	91	3363	84	220.565	0.480	1.583	1.574	0.983	0.991
CMU_2 / p-VCUBE_5	175	410	16	394	159	25.841	0.909	1.038	1.574	0.983	0.991
CMU_2 / p-VCUBE_6	175	27465	139	27326	36	1792.195	0.206	9.167	24.201	0.851	0.972
CMU_2 / p-VCUBE_7	175	11811	144	11667	31	765.189	0.177	4.003	0.262	0.989	0.990
CMU_2 / p-VCUBE_8	175	3721	94	3627	81	237.879	0.463	1.652	29.907	0.834	0.984
CMU_2 / p-VCUBE_9	175	551	26	525	149	34.432	0.851	1.024	0.262	0.989	0.990
INRIA-WILLOW_3 / p-SYS_1	175	33637	152	33485	23	2196.138	0.131	11.112	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_2	175	7729	92	7637	83	500.878	0.474	2.979	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_3	175	33637	149	33488	26	2196.334	0.149	11.130	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_4	175	7729	90	7639	85	501.009	0.486	2.991	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_5	175	33637	152	33485	23	2196.138	0.131	11.112	0.066	1.000	1.000
INRIA-WILLOW_3 / p-SYS_6	175	7729	91	7638	84	500.944	0.480	2.985	0.066	1.000	1.000
IPG-BJTU_5 / p-SYS_1	175	64	[illegible]	55	166	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PKU-IDM_5 / p-eSur_1	175	147	4	143	171	9.379	0.977	1.024	0.066	1.000	1.000
PKU-IDM_5 / p-eSur_2	175	158	4	154	171	10.100	0.977	1.028	2.296	0.983	0.994
PKU-IDM_5 / p-eSur_3	175	821	3	98	172	6.427	0.983	1.015	1.640	0.989	0.997
PKU-IDM_5 / p-eSur_4	175	925	6	71	169	4.657	0.966	0.989	4.788	0.960	0.984
TJU_2 / p-TJUMM_1	175	22882	146	22736	29	1491.157	0.166	7.622	20.725	0.880	0.984
TJU_2 / p-TJUMM_2	175	18808	149	18659	26	1223.764	0.149	6.267	0.066	0.994	0.995
TJU_2 / p-TJUMM_3	175	16152	144	16008	31	1049.896	0.177	5.427	0.066	0.994	0.995
TJU_2 / p-TJUMM_4	175	13482	142	13340	33	874.913	0.189	4.563	0.066	0.994	0.995
TJU_2 / p-TJUMM_5	175	10691	137	10554	38	692.192	0.217	3.678	0.066	0.994	0.995
TJU_2 / p-TJUMM_6	175	8162	123	8039	52	527.244	0.297	2.933	24.660	0.869	0.992
TJU_2 / p-TJUMM_7	175	5890	113	5777	62	378.889	0.354	2.249	28.005	0.846	0.986
TJU_2 / p-TJUMM_8	175	3672	86	3586	89	235.190	0.509	1.684	29.645	0.834	0.983



Summary

- Our participation in TRECVID-ED 2010
  - Submitted 4 event detection results
  - 3 of them obtain improvements over the best results of last year, especially on correct detection rate
  - Still have a much room for performance improvement!

- Making progress towards correct directions
  - *Selective eigenbackgrounds* to enable more effective foreground object extraction
  - *Multi-Pose Learning* for head-shoulder detection to address the data confusion problem
  - *Sequence Learning* for event detection: SVM-HMM by modeling the activity as sequence structure and exploring dynamics of the pattern within an event.

THANKS

Member: Yonghong Tian [a], Yaowei Wang [a] , Lei Qing [c]
Kaihua Jiang [b], Zhipeng Hu [a], Zhongwei Chen [c], Guochen Jia [a],
Ten Xu [a], Qiong Hu [c], Qiong Hu [c], Guangcheng Zhang [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