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

Empowered by Innovation NEC

PKU-NEC@TRECvid SED 2011: Sequence-Based Event Detection in Surveillance Video

Yonghong Tian[1] , Yaowei Wang[1,3] and Wei Zeng[2]

[1]National Engineering Laboratory for Video Technology, School of EE & CS, Peking University

[2] NEC Laboratories, China

[3] Department of Electronic Engineering, Beijing Institute of Technology

Outline

- Our System and Solutions @ 2011
  - Detection and Tracking
  - Pair-wise Event Detection
    - PeopleMeet, Embrace, PeopleSplitup
  - Action-Like Event Detection
    - ObjectPut, Pointing
- Summarization on Three Years' Experience of TrecVID SED
  - Our Participation Summarization
  - Revisit the Challenging Problems
  - Success and Lessons

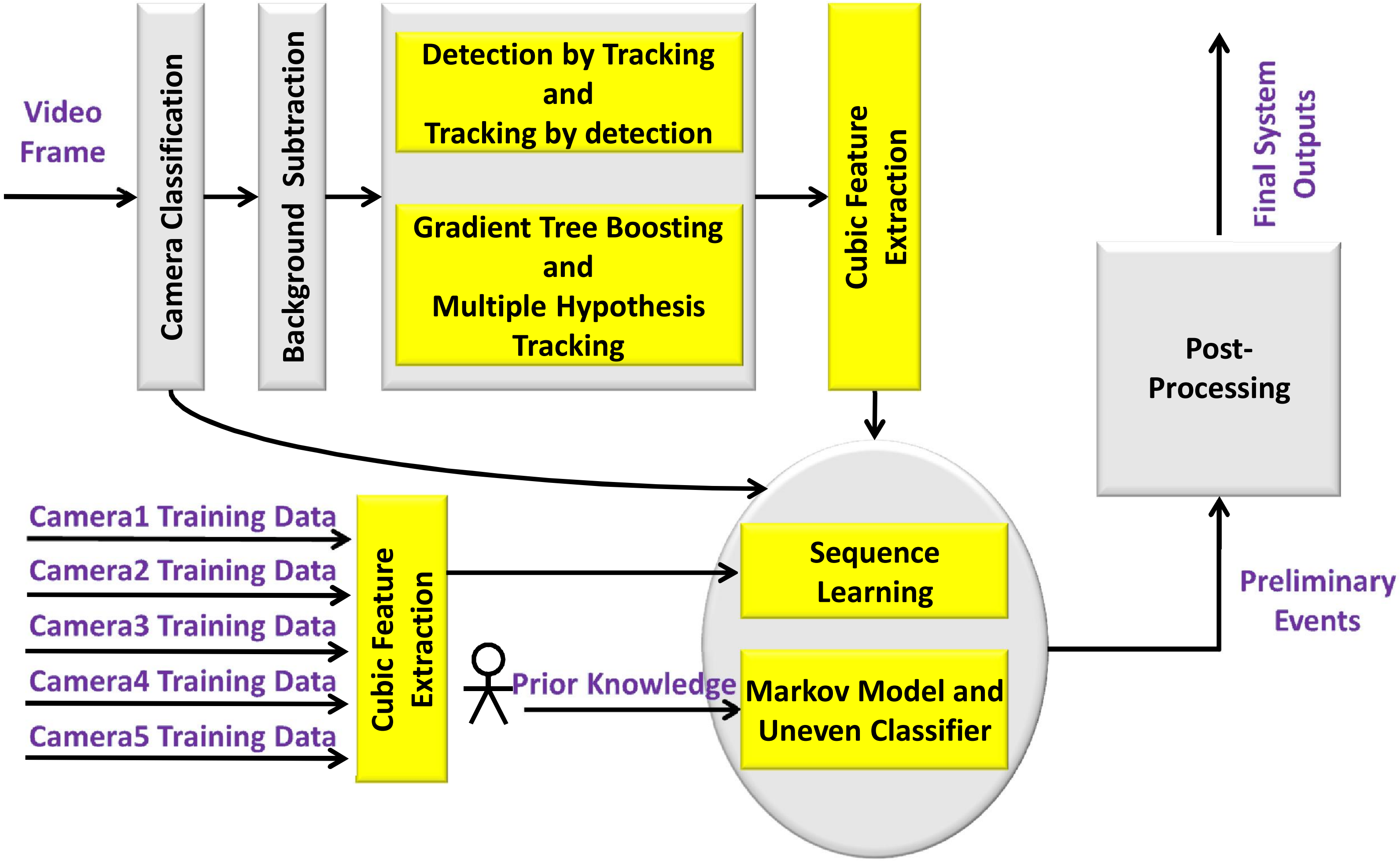
Acknowledgements

- Financial Support by NEC Lab China and NSFC
- Support and Advising
  - Prof. Wen Gao, and Prof. Tiejun Huang
  - Dr. Jun Du, and Mr. Atsushi Kashitani
- NEC Team
  - Wei Zeng, Hongming Zhang
  - Shaopeng Tang, Feng Wang, Guoyi Liu, Guangyu Zhu
- PKU Team
  - Yonghong Tian, Yaowei Wang
  - Xiaoyu Fang, Chi Su, Teng Xu, Ziwei Xia, and Peixi Peng

Empowered by Innovation

Our System and Solutions @ 2011

Framework of Our System



Empowered by Innovation NEC

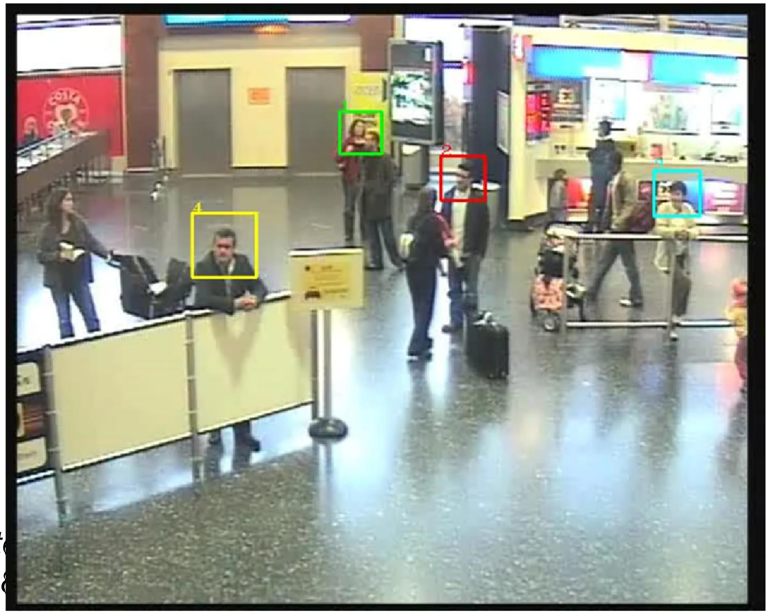
What are Key Points?

- Head-Shoulder Detection and Tracking
  - Detection-by-tracking and tracking-by-detection (By PKU Team)
  - Gradient Tree Boosting and Multiple Hypothesis Tracking (By NEC Team)
- Pair-wise Event Detection
  - Cubic Feature Extraction
  - Sequence Discriminant Learning using SVM^{DTAK}
- Action-like Event Detection
  - Markov chain based event modeling
  - Uneven SVM classifier

Our Solution (1): Detection &Tracking by PKU Team

- Motivation
  - Detection is not an isolated task!
  - Event detection needs an optimal output by integrating detect and tracking as one task.
- Detection-by-Tracking
  - Good Detection → Good Tracking?
  - Relatively good detection results in last year's system

	Cam1	Cam2	Cam3	Cam5
Precision	0.796	0.560	0.429	0.468
Recall	0.539	0.773	0.667	0.757
F1	**0.6429**	**0.6495**	**0.5222**	**0.5783**

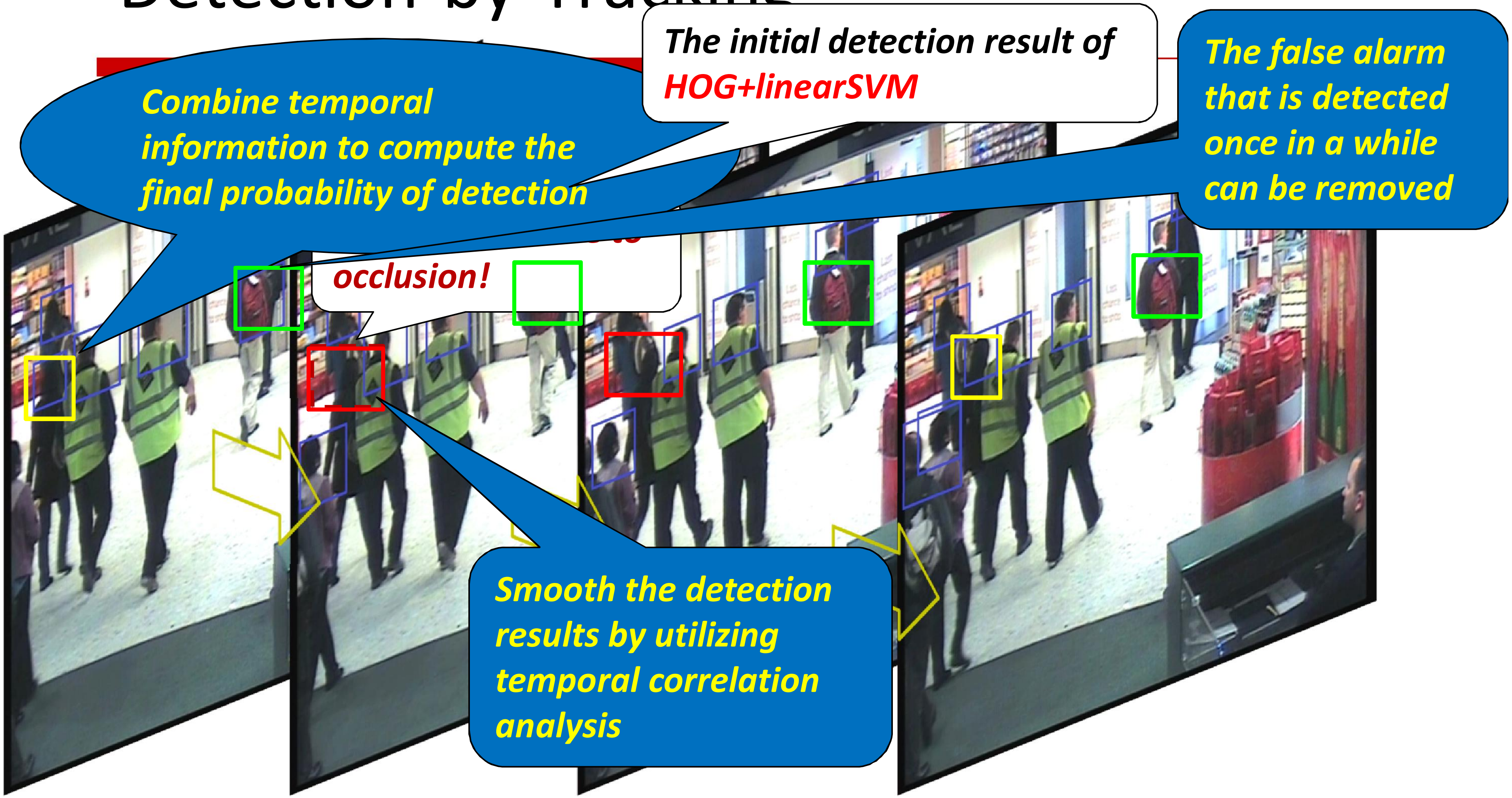


  - BUT the tracking……have many ID switches and drifts!

M. Andriluka, S. Roth, B. Schiele. People-tracking-by-detection and people-det… Conference on Computer Vision and Pattern Recognition (CVPR), Page(s): 1–…

Empowered by Innovation NEC

Detection-by-Tracking



- Combine the temporal information **like a tracker manner**
  - Confidence of HOG + linSVM detector
  - Appearance similarity
  - Location and scale similarity

Empowered by Innovation NEC

Detection-by-Tracking: Results

- On a labeled TRECVID 2008 corpus

Cam1			Cam2		
Recall	Precision	F-score	Recall	Precision	F-score
0.557	0.848	**0.6724**	0.372	0.785	**0.5048**
Cam3			Cam5		
0.423	0.756	**0.5425**	0.318	0.775	**0.4510**

Empowered by Innovation NEC

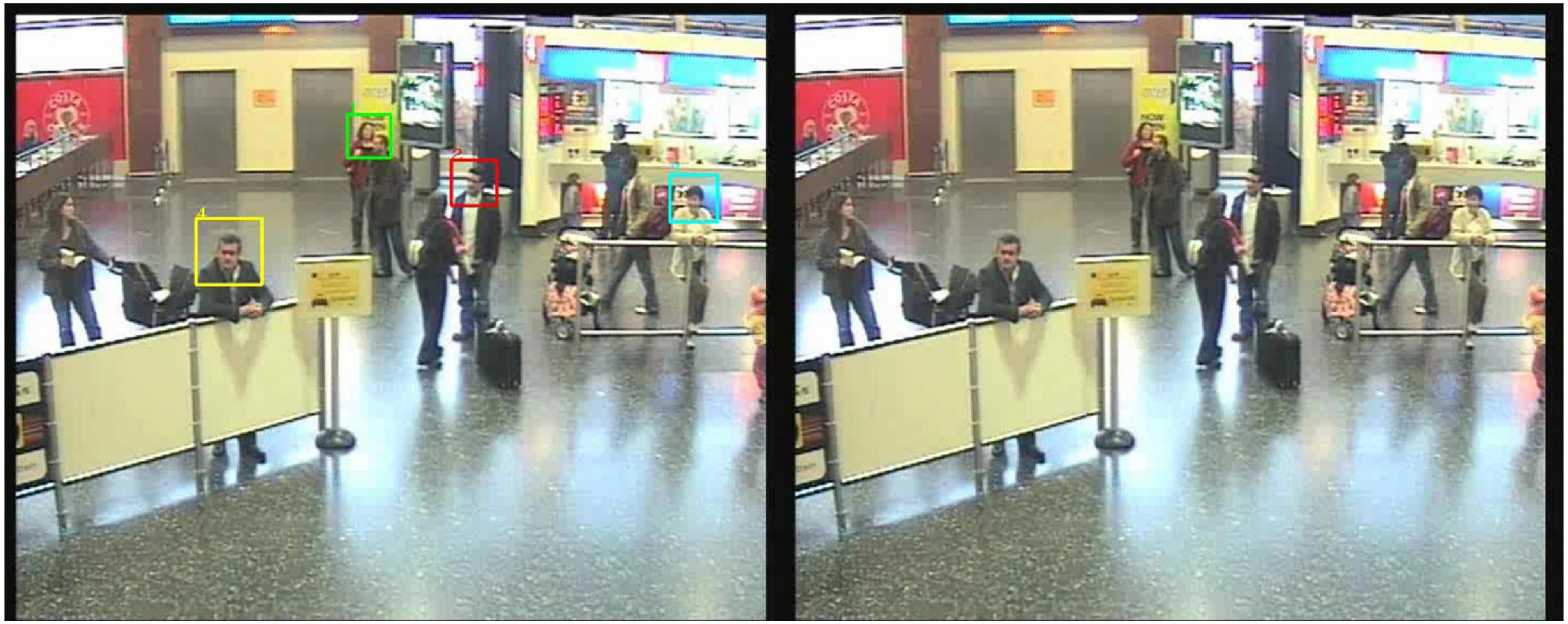
Our Solution (1): Detection &Tracking by PKU Team

- Motivation
  - **How to reduce ID switches and drifts?**
    - Complex human interactions
    - Heavy occlusion
- Tracking by detection
  - Link detection responses to trajectories by global optimization based on position, size and appearance similarities
  - Combine object detectors and particle filtering results in the algorithm [Breitenstein, 2010]

Michael D. Breitenstein, Fabian Reichlin, Bastian Leibe, Esther Koller-Meier, Luc Van Gool. Online Multi-Person Tracking-by-Detection from a Single, Uncalibrated Camera. PAMI, 2010.

Tracking-by-Detection: Results

	Camera1	MOTA	MOTP	Miss	FA	ID Switch
Camera 1	Last Year	0.321	0.591	0.510	0.134	0.035
	This Year	0.364	0.567	0.472	0.154	0.010
Camera 2	Last Year	-0.135	0.599	0.791	0.317	0.027
	This Year	0.213	0.607	0.644	0.132	0.011
Camera 3	Last Year	0.022	0.571	0.652	0.293	0.033
	This Year	0.271	0.591	0.667	0.050	0.010
Camera 4	Last Year	-0.002	0.602	0.537	0.440	0.025
	This Year	0.170	0.589	0.731	0.089	0.009



Empowered by Innovation NEC

Our Solution (2): Detection &Tracking by NEC Team

Detection with Gradient Tree Boosting

- Use cascade gradient boosting [Friedman 01] as a learning framework to combine decision trees to form a simple and highly robust object classifier.
- Instead of SVM, we use decision tree algorithm as weak classifier.

Experimental Results

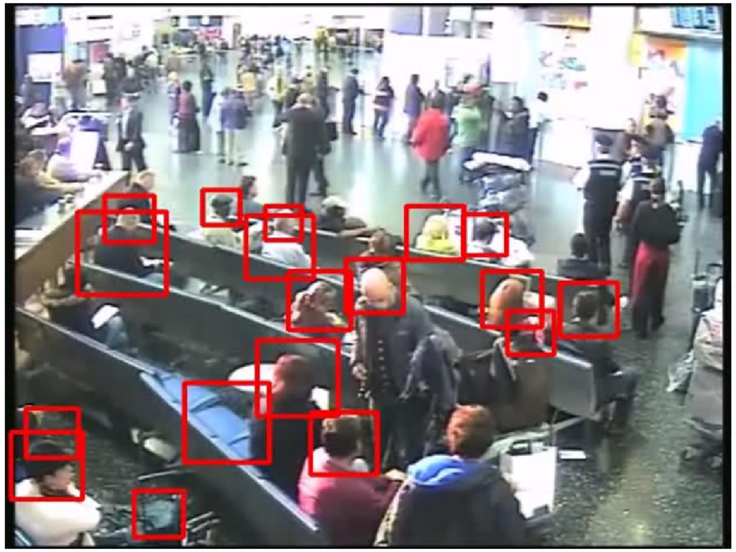
- On a labeled TRECVID 2008 corpus

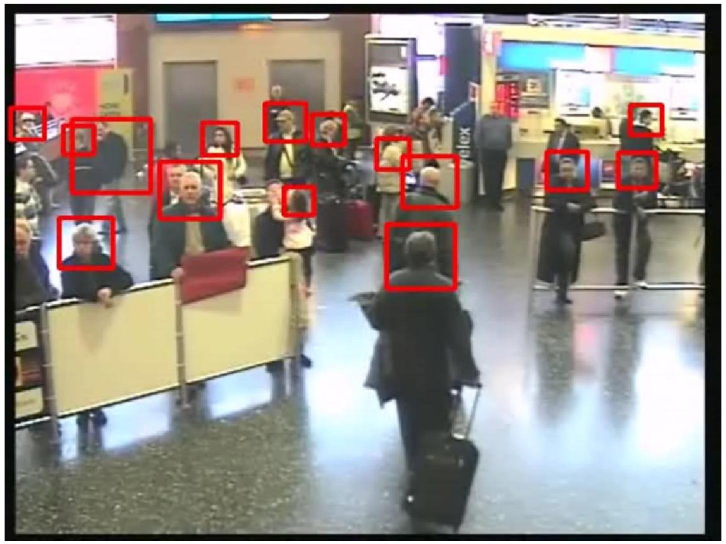
Cam1			Cam2		
Recall	Precision	F-score	Recall	Precision	F-score
0.553	0.803	**0.6550**	0.356	0.727	**0.4780**
Cam3			**Cam5**		
Recall	Precision	F-score	Recall	Precision	F-score
0.294	0.801	**0.4301**	0.271	0.732	**0.3755**

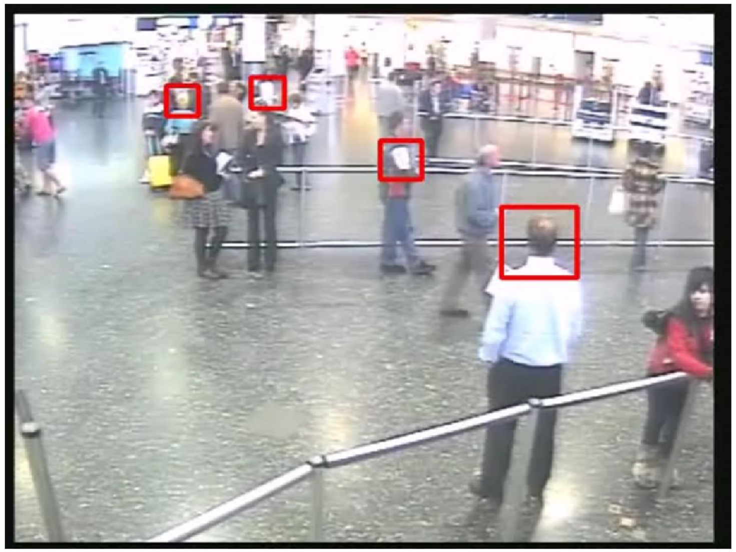
[Friedman 01] J. Friedman. Greedy Function Approximation: A Gradient Boosting Machine. Ann. Statist. 29(5), 2001, 1189-1232.

Demo for Gradient Tree Boosting


Cam 1

Cam 2



Cam 3

Cam 5


MHT Tracking

- In order to track multiple objects in TRECVID video, we adopt Multiple Hypothesis Tracking (MHT) [Cox 96] Method.

	MOTA	MOTP	Miss	FA	ID Switch
Camera1	0.368	0.571	0.486	0.134	0.012
Camera2	0.151	0.601	0.680	0.160	0.009
Camera3	0.198	0.583	0.746	0.051	0.005
Camera5	0.168	0.591	0.737	0.088	0.008

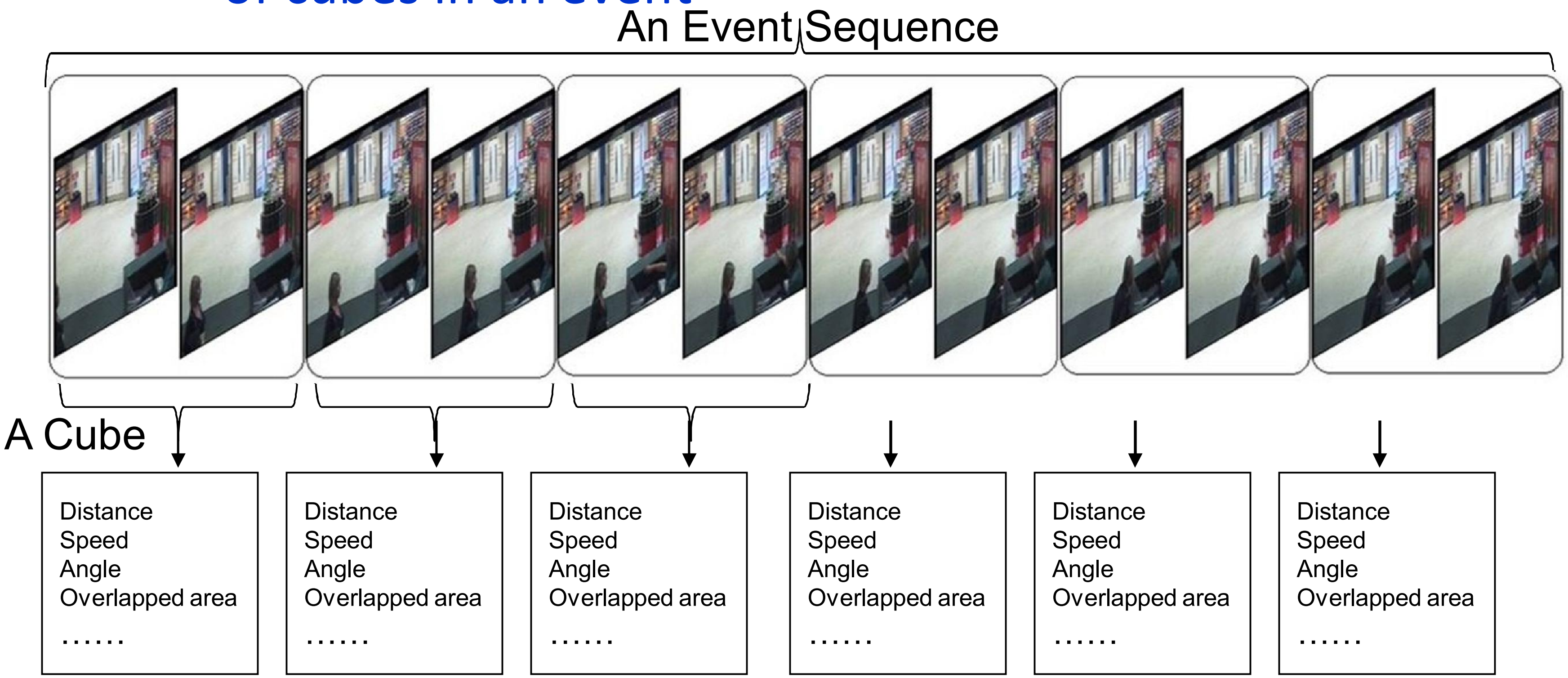
[Cox96] I.J. Cox, S.L. Hingorani, An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking, PAMI, 18(2), 138 – 150, 1996

Our Solution (3):

Empowered by Innovation NEC

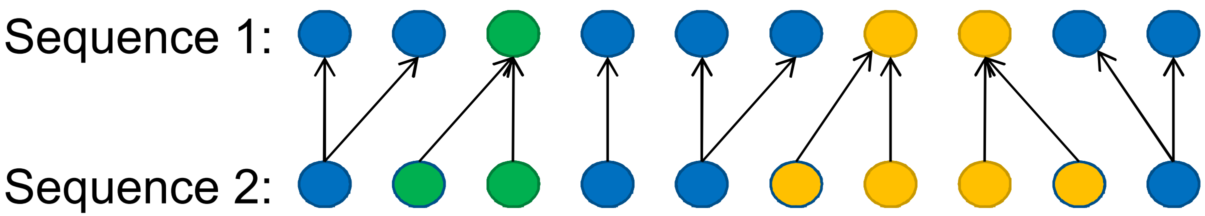
Sequence Learning for Pair-wise Event Detection

- Event analysis based on sequence learning
  - Model the activity as sequence structure and consider the information in and between frames
  - Cubic Feature: Fixed cube length and variable numbers of cubes in an event



Pair-wise Event Detection

- SVM over Dynamic Time Alignment Kernel
  - Dynamic time wrapping: Find an optimal path φ to minimize the distance of two sequences.



They have the same pattern using Dynamic Time Alignment kernel !!!

$$K(X,Y) = D_{\phi}(X,Y) = \frac{1}{N}\sum_{n=1}^{N} k(x_{\phi_{X(n)}}, y_{\phi_{Y(n)}})$$

Empowered by Innovation

Experimental Results

- Evaluation on 10 hours data from TREVID-SED 2008 corpus
  - Based on detecting and tracking results
  - Compare with SVM and SVM^{HMM} approaches

event	#Ref		#Sys	#CorDet	#FA	#Miss	Min.DCR
PeopleMeet	298	★	54	7	47	291	1.000
		◇	29	2	27	296	1.007
		#	8	6	2	292	0.981
PeopleSplitUp	152	★	81	7	74	145	0.991
		◇	21	0	21	152	1.011
		#	164	23	141	129	0.919
Embrace	116	★	82	5	77	111	0.995
		◇	44	1	43	115	1.000
		#	7	3	4	113	0.976

**Without any post-processing*

★ is results of SVM^{HMM}
◇ is results of ordinary SVM
is results of SVM-DTAK

Obtain some performance improvement

Empowered by Innovation

Evaluation Results – PeopleMeet

EVENT : PeopleMeet	Inputs		Actual Decision DCR Analysis				Minimum DCR Analysis
	#Targ	#Sys	#CorDet	#FA	#Miss	DCR	DCR
PKUNEC_6 p-eSur_3	449	2382	24	108	425	**0.982**	0.9777
CMU_8 p-SYS_1	449	381	45	336	404	1.01	0.9724
TokyoTech-Canon_1 p-HOG-SVM_1	449	3949	8	140	441	1.0281	1.0003
BUPT-MCPRL_7 p-baseline_1	449	886	55	831	394	1.15	1.0119
TJUT-TJU_10 p-VCUBE_7	449	3491	140	3351	309	1.7871	0.9848
IRDS-CASIA_5 p-baseline_1	449	8262	294	7968	155	2.9581	0.9997

Empowered by Innovation NEC

Evaluation Results - Embrace

EVENT : Embrace	Inputs		Actual Decision DCR Analysis				Minimum DCR Analysis
	#Targ	#Sys	#CorDet	#FA	#Miss	DCR	DCR
CMU_8 p-SYS_1	175	715	58	657	117	0.884	0.8658
PKUNEC_6 p-eSur_3	175	5234	15	102	160	0.9477	0.9453
NHKSTRL_3 p-NHK-SYS1_3	175	3869	31	804	144	1.0865	1.0003
CRIM_4 p-baseline_1	175	1205	25	1180	150	1.2441	1.0003
BUPT-MCPRL_7 p-baseline_1	175	3382	74	3308	101	1.6619	1.0008
TJUT-TJU_10 p-VCUBE_7	175	4623	104	4519	71	1.8876	0.9934
IRDS-CASIA_5 p-baseline_1	175	9693	152	9541	23	3.2602	1.0003

Empowered by Innovation NEC

Evaluation Results – PeopleSplitUp

EVENT : PeopleSplitUp	**Inputs**		**Actual Decision DCR Analysis**				**Minimum DCR Analysis**
	#Targ	**#Sys**	**#CorDet**	**#FA**	**#Miss**	**DCR**	**DCR**
TokyoTech-Canon_1 p-HOG-SVM_1	187	2595	51	557	136	0.9099	0.9066
BUPT-MCPRL_7 p-baseline_1	187	1009	59	950	128	0.996	0.8809
CMU_8 p-SYS_1	187	118	3	115	184	1.0217	1.0003
PKUNEC_6 p-eSur_3	187	2988	4	192	183	1.0416	1.0003
TJUT-TJU_10 p-VCUBE_7	187	436	13	423	174	1.0692	0.9901
IRDS-CASIA_5 p-baseline_1	187	4339	139	4200	48	1.634	0.9835

Empowered by Innovation NEC

Analysis of PeopleSplitUp

- The reason of SplitUp's low performance
  - Inconsistence of the evaluation parameter DeltaT between Task Webpage and Act. Used.
    - 10 → 0.5
  - Our mistakes: The event alignment is not accurate
    - The begin and end are not defined clearly
- Experimental results

event	#Ref		#Sys	#CorDet	#FA	#Miss	DCR
PeopleSplitUp	152	◇	21	0	21	152	1.011
		★	81	7	74	145	0.991
		#	164	23	141	129	0.919

**Without any post-processing*

◇ is results of ordinary SVM -— Used in 2009

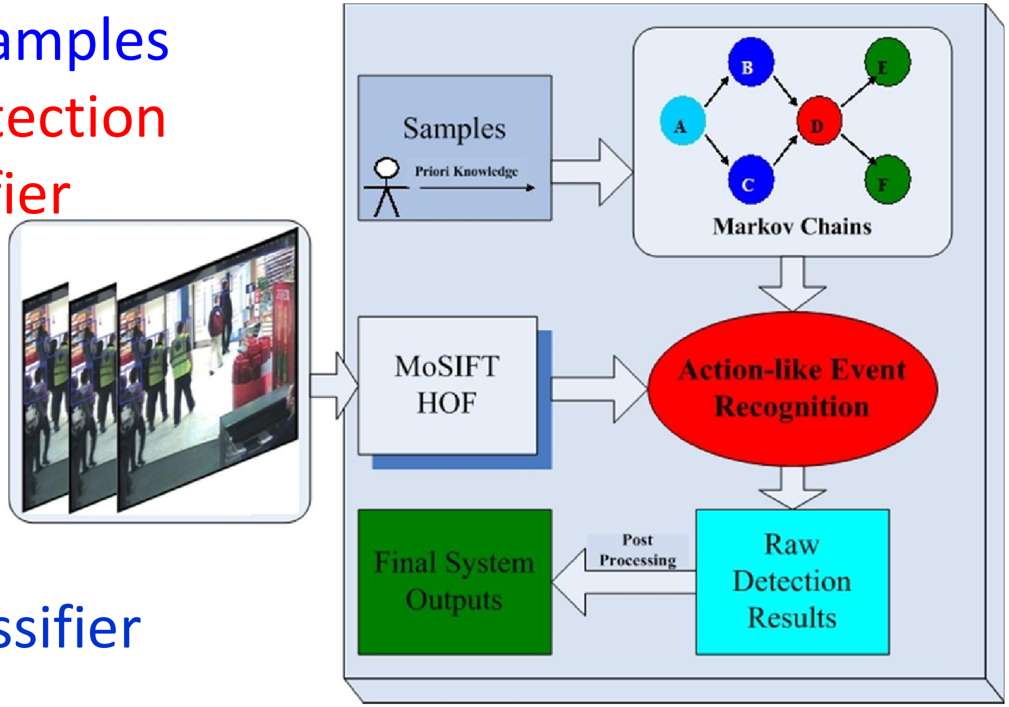
★ is results of SVM^{HMM} -— Used in 2010

is results of SVM-DTAK -— Used in 2011

Our Solution (4):
Uneven Classifier for Action-like Event Detection

Empowered by Innovation NEC

- Problem:
  - Few occurrences for each activity
  - Too many negative examples →Very few correct detection with the normal classifier
- Event detection with the uneven classifier
  - Modeling the activity with a Markov chain
  - Using uneven SVM classifier



Empowered by Innovation NEC

SVM with Uneven Margins

- The commonly used SVM model: Treats positive and negative training examples equally

$$\begin{aligned} \text{minimise}_{\mathbf{w},\, b,\, \xi} \quad & \langle \mathbf{w}, \mathbf{w} \rangle + C \sum_{i=1}^{l} \xi_i \\ \text{subject to} \quad & \langle \mathbf{w}, \mathbf{x}_i \rangle + \xi_i + b \geq 1 \quad \text{if } y_i = +1 \\ & \langle \mathbf{w}, \mathbf{x}_i \rangle - \xi_i + b \leq -1 \quad \text{if } y_i = -1 \\ & \xi_i \geq 0 \quad \text{for } i = 1, ..., m \end{aligned}$$

where C is the cost factor measures the cost of mistakenly classified examples in training set.

- SVM with Uneven Margins: Sets the positive margin be some larger than the negative margin.

$$\begin{aligned} \text{minimise}_{\mathbf{w},\, b,\, \xi} \quad & \langle \mathbf{w}, \mathbf{w} \rangle + C_\tau \sum_{i=1}^{l} \xi_i \\ \text{subject to} \quad & \langle \mathbf{w}, \mathbf{x}_i \rangle + \xi_i + b \geq 1 \quad \text{if } y_i = +1 \\ & \langle \mathbf{w}, \mathbf{x}_i \rangle - \xi_i + b \leq -\tau \quad \text{if } y_i = -1 \\ & \xi_i \geq 0 \quad \text{for } i = 1, ..., m \end{aligned}$$

Solve by the ordinary SVM →

$$\begin{aligned} \mathbf{w}_2^* &= \frac{1+\tau}{2}\mathbf{w}_1^* \\ b_2^* &= \frac{1+\tau}{2}b_1^* + \frac{1-\tau}{2} \\ \xi_2^* &= \frac{1+\tau}{2}\xi_1^* \end{aligned}$$

τ is the ratio of negative margin to positive margin of the classifier, $C_\tau = \frac{1+\tau}{2}C$

Y.Y. Li, J. Shawe-Taylor, The SVM With Uneven Margins snd Chinese Document Categorisation, PACLIC'03, 2003.

Empowered by Innovation NEC

Evaluation Results - ObjectPut

	Inputs		Actual Decision DCR Analysis				Minimum DCR Analysis
	#Targ	#Sys	#CorDet	#FA	#Miss	DCR	DCR
PKUNEC_6 p-eSur_3	621	50	8	41	613	1.0006	0.9983
CMU_8 p-SYS_1	621	58	1	57	620	1.0171	1.0003
NHKSTRL_3 p-NHK-SYS1_3	621	9216	10	552	611	1.1649	1.0003
TJUT-TJU_10 p-VCUBE_7	621	790	17	773	604	1.2261	1.0003
CRIM_4 p-baseline_1	621	2867	62	2805	559	1.82	1
BUPT-MCPRL_7 p-baseline_1	621	3643	111	3532	510	1.9795	1.0063
IRDS-CASIA_5 p-baseline_1	621	13746	343	13403	278	4.8429	0.9994

Empowered by Innovation NEC

Evaluation Results - Pointing

	Inputs		Actual Decision DCR Analysis				Minimum DCR Analysis
	#Targ	#Sys	#CorDet	#FA	#Miss	DCR	DCR
BJTU-SED_1 p-SYS_1	1063	88	36	37	1027	0.9783	0.973
PKUNEC_6 p-eSur_3	1063	2113	21	123	1042	1.0206	1.0032
NHKSTRL_3 p-NHK-SYS1_3	1063	13974	41	1237	1022	1.3671	1.0003
CMU_8 p-SYS_1	1063	2092	132	1960	931	1.5186	1.0001
TJUT-TJU_10 p-VCUBE_7	1063	2240	141	2099	922	1.5557	0.9994
BUPT-MCPRL_7 p-baseline_1	1063	4245	268	3977	795	2.0521	1.0003
IRDS-CASIA_5 p-baseline_1	1063	13733	654	13079	409	4.6737	1.0003
CRIM_4 p-baseline_1	1063	14089	582	13507	481	4.8818	1.0003

Summarization on Three Years’ Experience of TrecVID SED

Our Participations

2009

- PeopleMeet
- PeopleSplitUp
- Embrace
- ElevatorNoEntry
- PersonRuns

2010

- PeopleMeet
- PeopleSplitUp
- Embrace
- PersonRuns

2011

- PeopleMeet
- PeopleSplitUp
- Embrace
- ObjectPut
- Pointing

Collaborating with NEC Lab China!

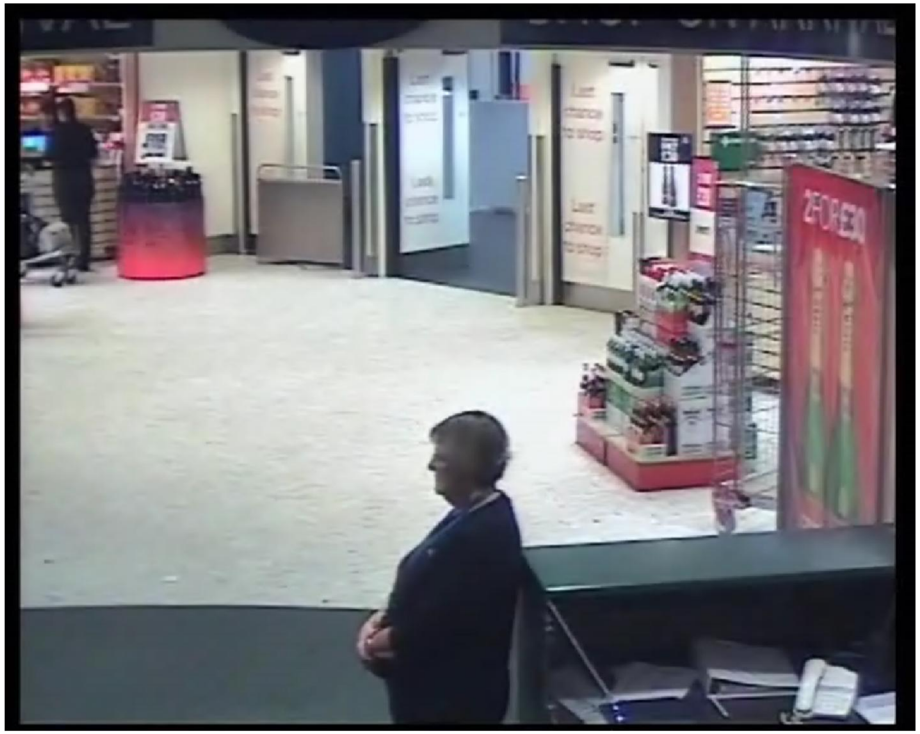
Empowered by Innovation NEC

Revisit: Challenges (1)

- No clear definition of begin and end of an event
- Examples:
  - **PeopleMeet Description:** One or more people walk up to one or more other people, stop, and some communication occurs.
    - Start Time: The first communication between members of two groups
    - End Time: The earliest time when the two groups are nearest to each other after the communica
  - **Problem:**
    - How to define groups ?
    - How to measure whether two groups are nearest?

Revisit: Challenges (2)

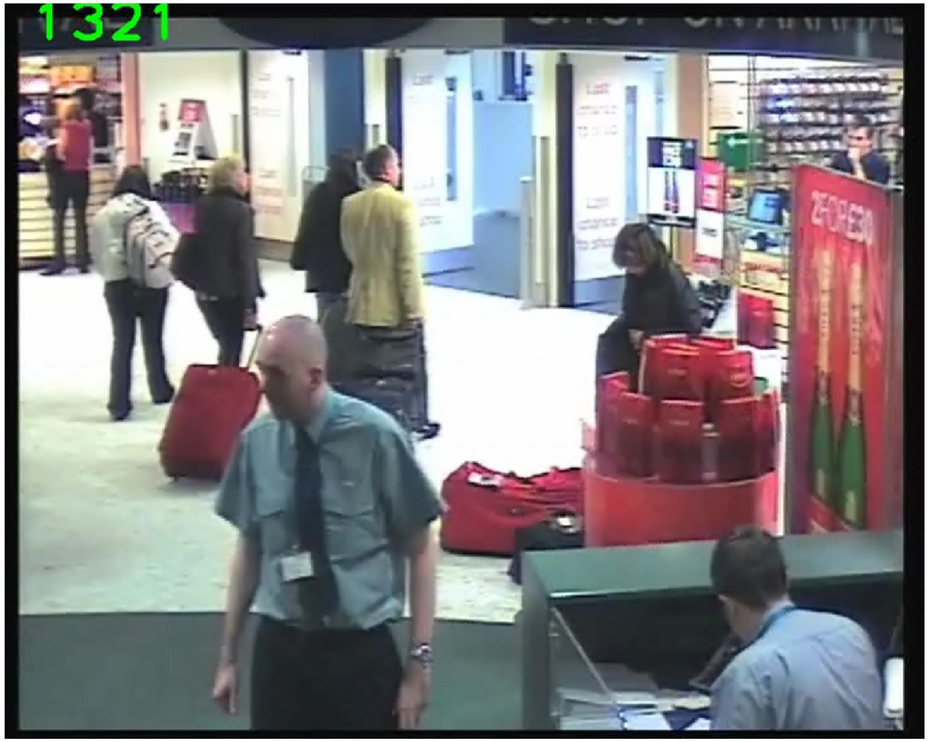
- Event's variance
  - For example: ObjectPut events are very different



Empowered by Innovation NEC

Revisit: Challenges (3)

- Event's similarity
  - [Pointing](#) VS [Arm Lift](#)

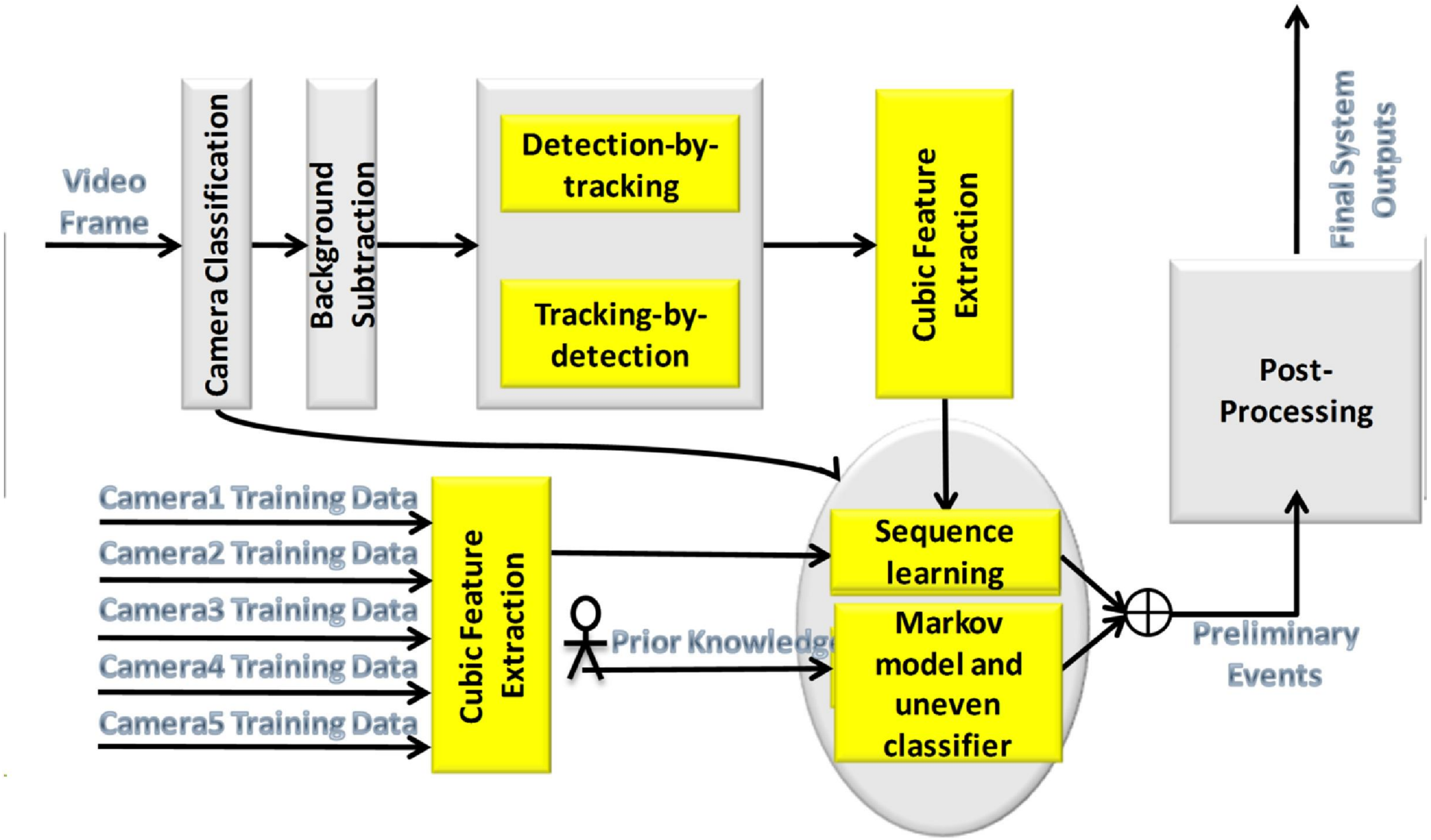


Developments of Our Systems

2011: Detect by temporal feature and sequence learning

2009: Detect by frame feature and normal learning method

2010: Detect by frame feature and SVM-HMM



Improvement of Results

Results Comparison

CorDet greatly Increased

Better than do nothing

PeopleMeet	#Ref	#Sys	#CorDet	#FA	#Miss	Act.DCR
2011	449	2382	**24**	108	425	**0.982**
2010	449	156	**12**	144	437	**1.02**
2009	449	125	**7**	118	442	**1.023**
Embrace						
2011	175	5234	**15**	102	160	**0.9477**
2010	175	925	**6**	71	169	**0.989**
2009	175	80	**1**	79	174	**1.020**
PeopleSplitUp						
2011	187	2988	**4**	192	183	**1.0416**
2010	187	167	**16**	136	171	**0.959**
2009	187	198	**7**	191	180	**1.025**

Summary: Success

- Making progress towards correct directions
  - Detection + Tracking:
    Boosting
    →Multiple Pose Learning + Multiple Instance Learning
    →Detection-by-tracking + Tracking-by-detection
  - Feature:
    Frame-based
    → Temporal Cubic Feature
  - Event Learning methods:
    Normal SVM + Automata
    → SVM-HMM
    → SVM-DTAK + Uneven Classifier

Summary: Lessons

- For detection and tracking, there are much room for improvement.
  - The dataset is too complex for detection and tracking algorithms on a single, uncalibrated camera!
  - Crowded scene detection and tracking is still a challenging problem.
- The event detection is far from practical applications.
  - Unclear event definition will *mislead* the development of algorithms.
  - Have to consider the uneven distribution of abnormal events

شكراً

ขอบคุณ

DMnvwd

Gracias

Dankie

Obrigado!

ありがとう！

WAD MAHAD
SAN TAHAY

Ευχαριστώ

Vielen Dank

谢谢！

THANK YOU

شكريا

감사합니다

Asante

متشكرم

Merci

Teşekkürler

Hvala

Grazie

Bedankt

Köszönettel

Díky

go raibh maith agaibh

GADDA GUEY

Urakoze