

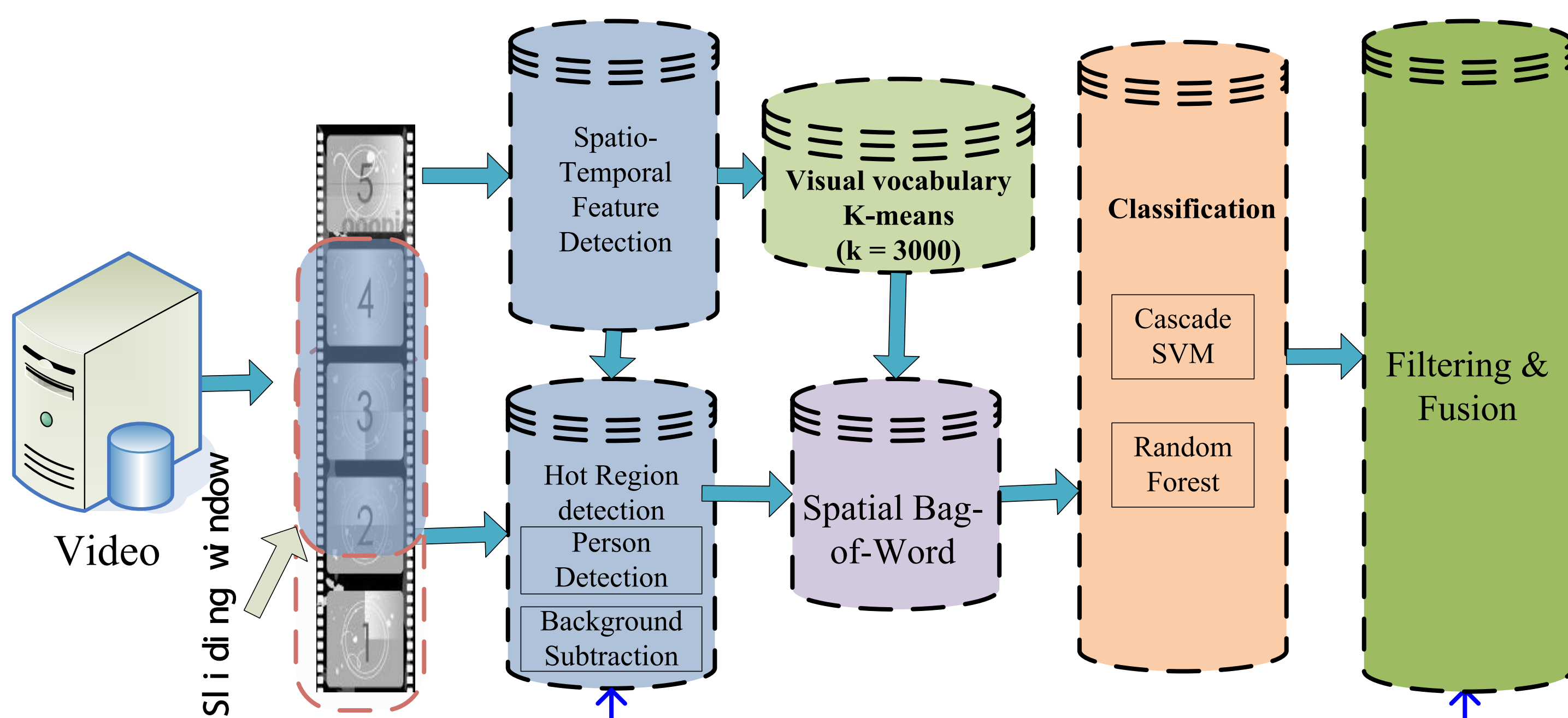
INFORMEDIA@TRECVID 2011: SURVEILLANCE EVENT DETECTION

Longfei Zhang, Lu Jiang, Lei Bao, Shohei Takahashi, Yuanpeng Li, Alexander Hauptmann(CMU), Babu Pillai (Intel)

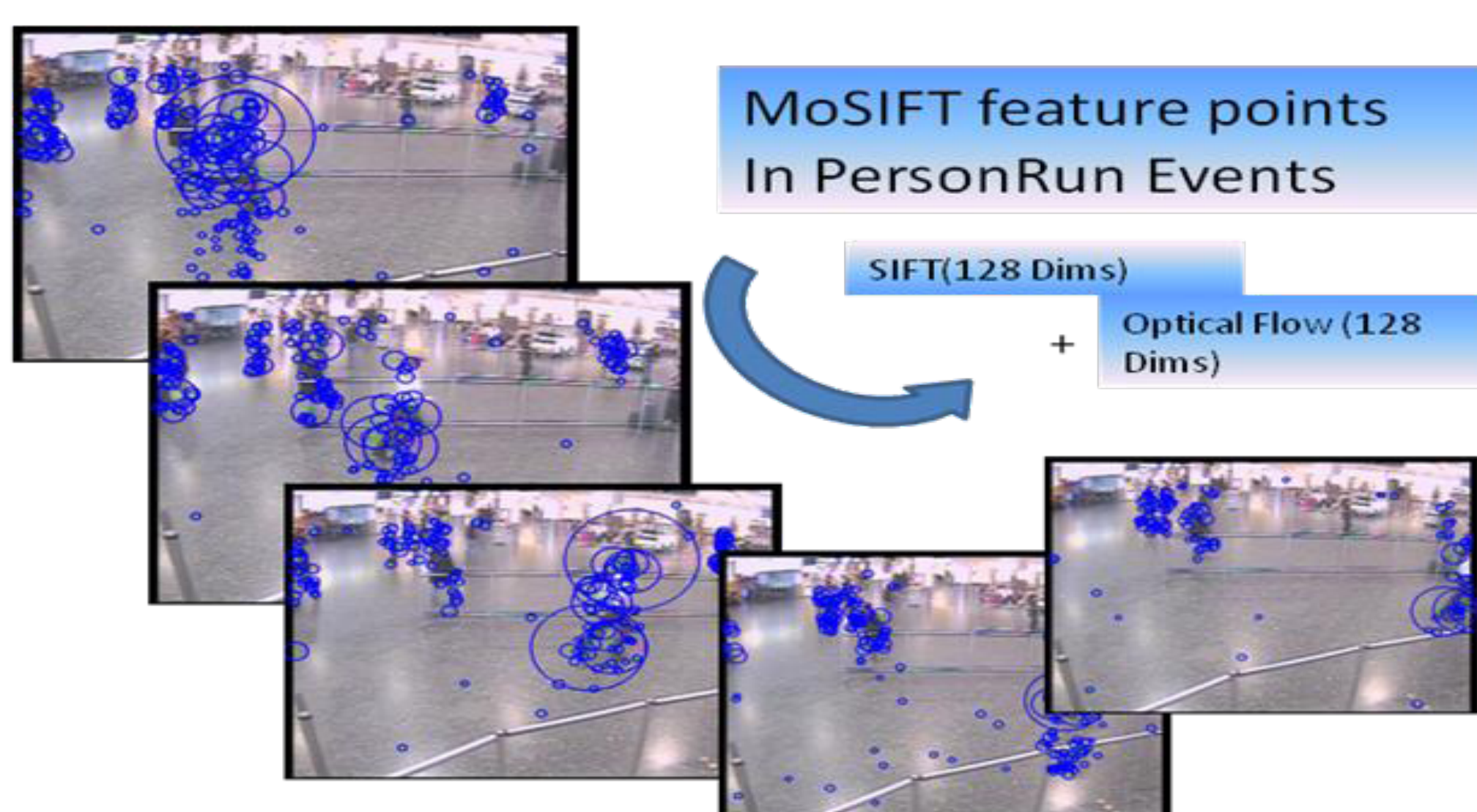
INTRODUCTION

- Surveillance Event Dataset:
 - 144-hour surveillance video dataset recorded in London Gatwick Airport through 5 different cameras (fixed cameras).
- Challenges:
 - Large-scale real-world video data analysis.
 - Difficult human actions detection from observation.
 - Small training data set & imbalance data.

FRAMEWORK: AUGMENTED BOOSTED CASCADE

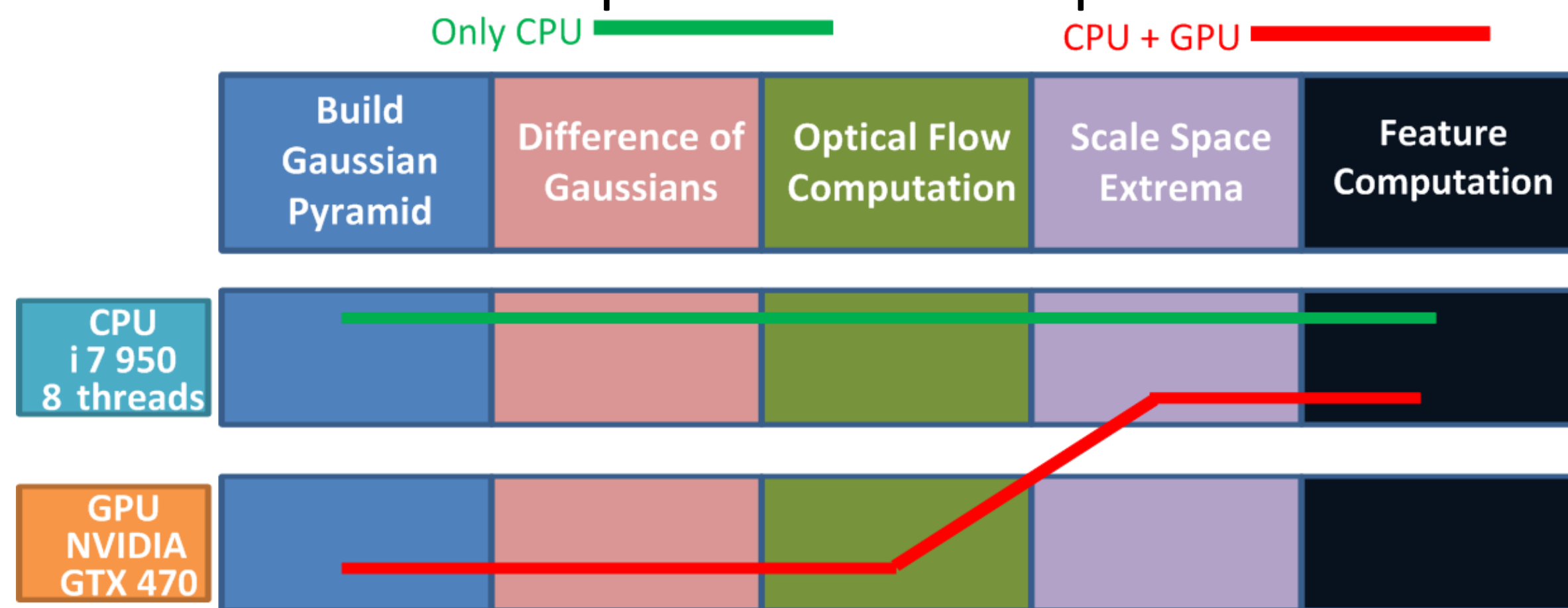


MOSIFT BASED ACTION RECOGNITION

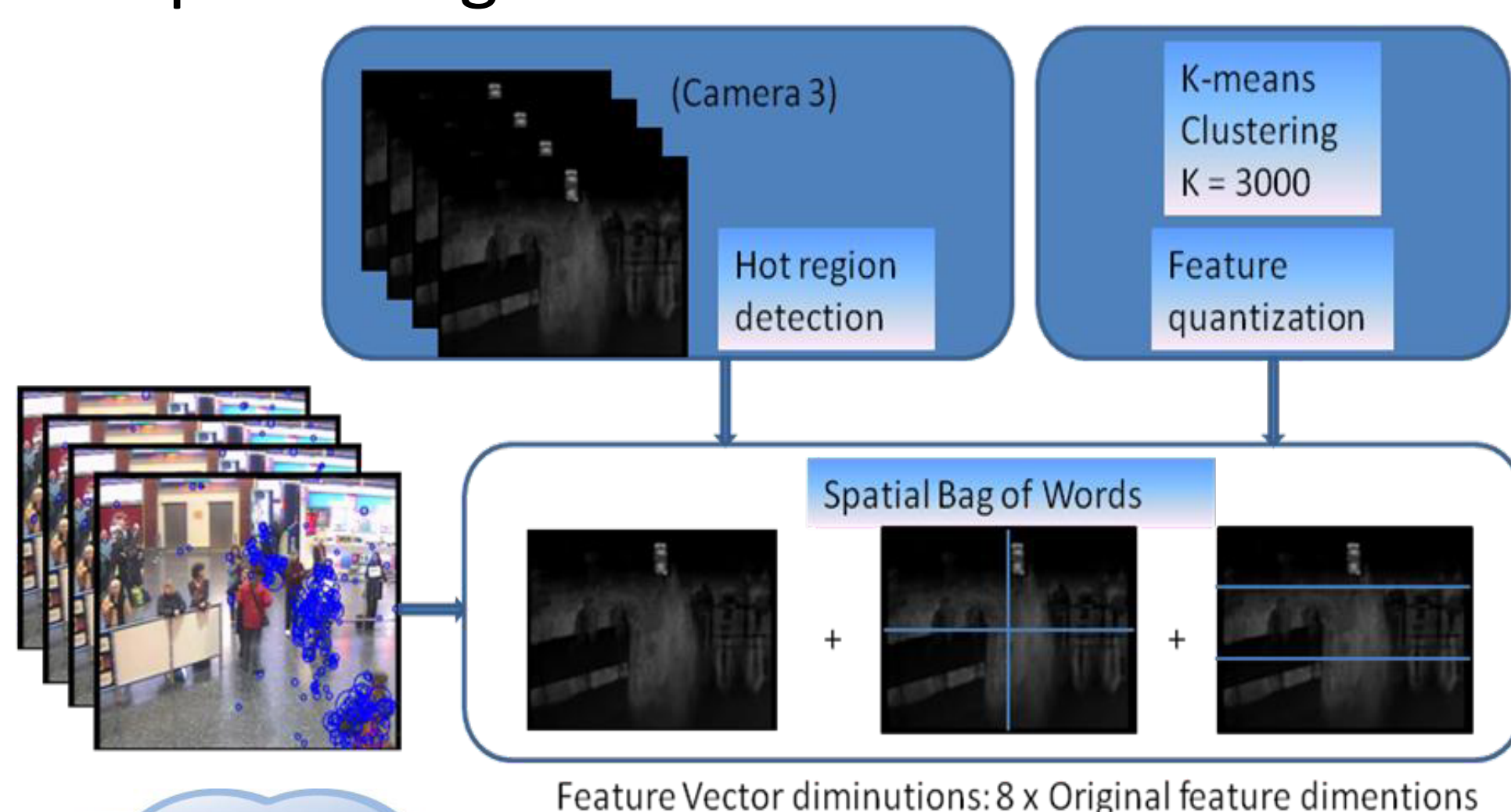


- Takes a pair of video frames to find spatio-temporal interest points.
- Detecting SIFT points and using optical flow to match the scale of the SIFT points.

- MoSIFT on GPU: Computes first three phases of MoSIFT on GPU.



- Spatial Bag of Words



- Each frame is divided into rectangular tiles or grids.
- The Spatial BoW facilitates the classifier distinguish events happen in different part of the frame thus encodes spatial information in features.

RESULTS IN NIST TRECVID COMPETITION

- Evaluation Criteria- DCR(Detection Cost Rate)

$$DCR = P_{Miss} + 0.005 \times P_{FA}$$

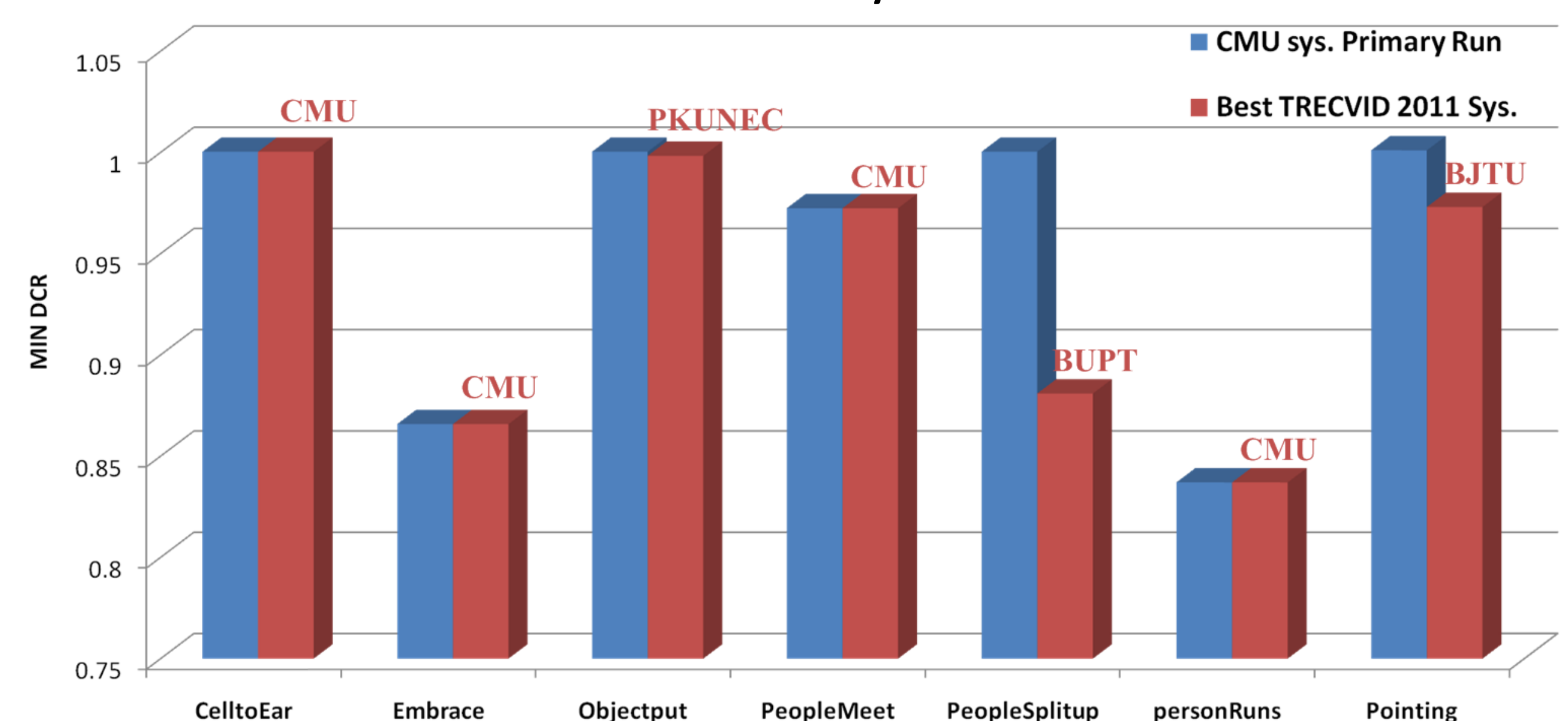
P_{Miss} is the number of missed detections divides the number of observations. P_{FA} is the number of false alarms divides the total duration of the video segments in hours.

PERFORMANCE

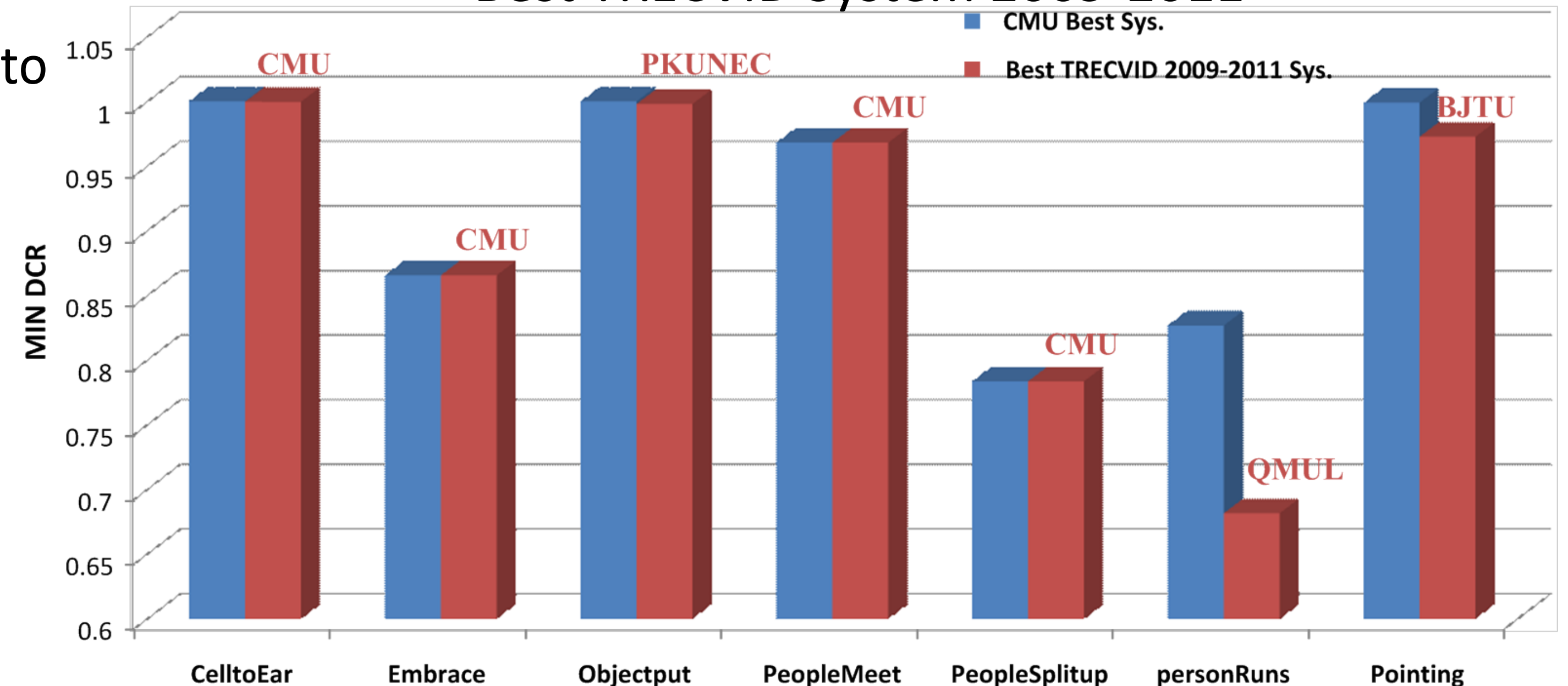
Event	Primary Run					Best Runs	
	MinDCR	ADCR	#CorDet	#FA	#Miss	Actual DCR	MinDCR
CellToEar	1.0003	1.0365	1	127	193	1.0111	1.0003
Embrace	0.8658	0.8840	58	657	117	0.8769	0.8658
ObjectPut	1.0003	1.0171	0	57	620	1.0023	1.0003
PeopleMeet	0.9724	1.0100	45	336	404	1.01	0.9684
PeopleSplitUp	1.0003	1.0217	3	115	184	0.9569	0.7838
PersonRuns	0.8370	0.8924	26	413	81	0.8577	0.827
Pointing	1.0001	1.5186	132	1960	931	1.0201	0.9994

COMPARISON OF BEST SYSTEMS

Best TRECVID System 2011



Best TRECVID System 2009-2011



MOSIFT ON GPU IN REAL TIME

	MoSIFT average seconds taken per phase (749 frames) No doubling of Image		
	Only CPU i7 1 thread	Only CPU i7 8 threads	i7 8 threads and GTX 470 GPU
Build Gaussian Pyramid	0.0538	0.0335	0.0194
Difference of Gaussians	0.0035	0.0042	0.0023
Optical Flow Computation	0.1790	0.0621	0.0156
Scale Space Extrema	0.0212	0.0210	0.0107
Feature Computation	0.0047	0.0017	0.0029
Total time per frame (seconds)	0.2622	0.1225	0.0509

- Only CPU: 8 frames per second.
- CPU + GPU: 20 frames per second (real-time!!)