

The PASCAL Visual Object Classes Challenge 2010 (VOC2010)

Part 4 – Action Classification Taster Challenge

Mark Everingham

Luc Van Gool

Ivan Laptev

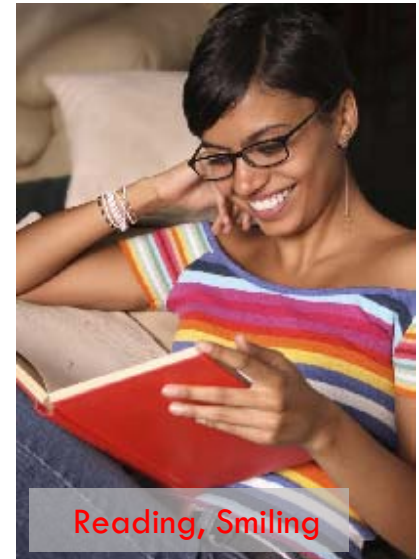
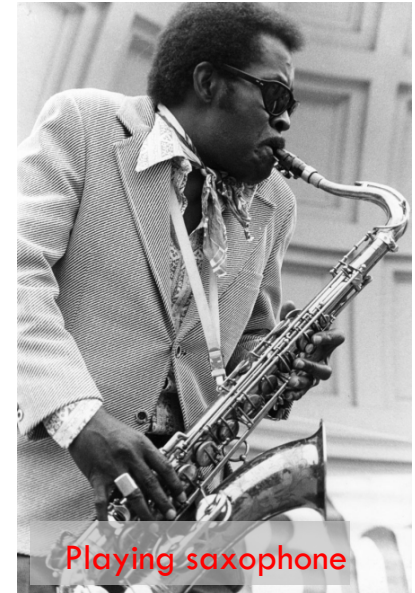
Chris Williams

John Winn

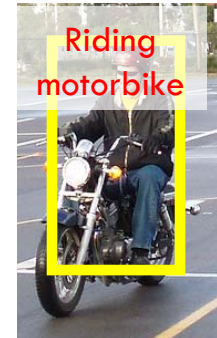
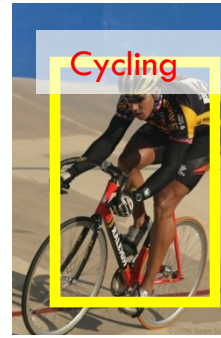
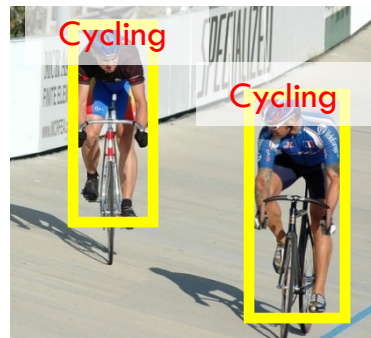
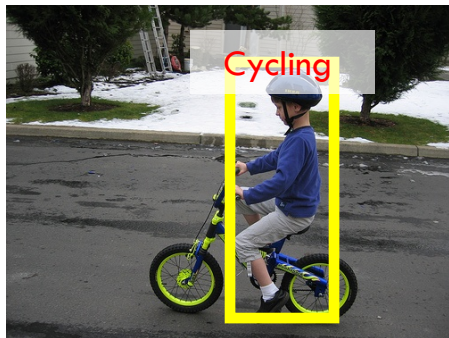
Andrew Zisserman



Motivation I: Actions describe many images



Motivation II: Actions in PASCAL VOC2009

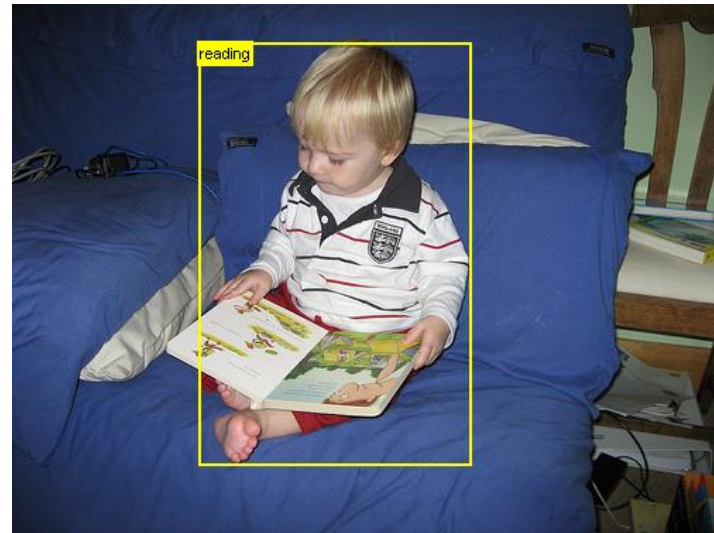


Action Classification Taster Challenge

- Given the bounding box of a person, predict whether they are performing a given action



Playing Instrument?



Reading?

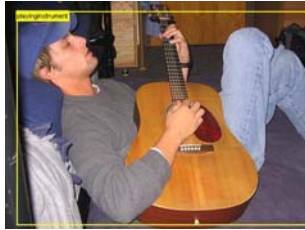
- Encourage research on **still-image** activity recognition: more detailed understanding of image

Nine Action Classes

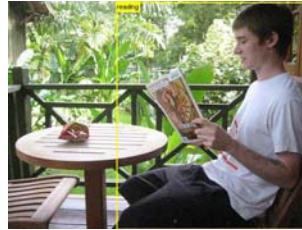
Phoning



Playing Instrument



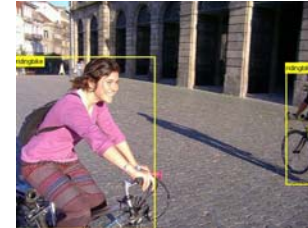
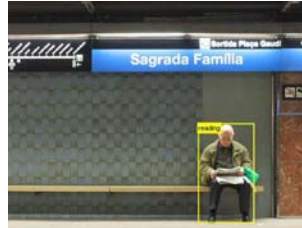
Reading



Riding Bike



Riding Horse



Running



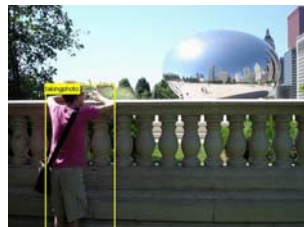
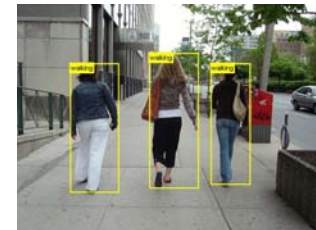
Taking Photo



Using Computer



Walking



Dataset Statistics

- Images collected from flickr using action queries
 - Disjoint to main challenge dataset

	Training	Testing
Images	454	454
Objects	608	613

- 50-100 training objects per class
- Only subset of people are annotated (bounding box + action)
- All people in dataset are labelled with exactly one action class
 - In future actions will not be mutually exclusive (or complete?)

Methods

- Comp9 (Train on VOC data): 11 Methods, 8 Groups
 - Image classification within bounding box
 - SVM, bag of words/spatial pyramid
 - Multiple features: SIFT, PHOG, color SIFT, etc.
 - Context (image, bounding box, neighbouring region)
 - Classification of multiple figure-ground segmentations
 - Combined image classification and part-based detection
- Comp10 (Train on own data): 1 Method
 - Poselets, object context

AP by Class/Method

Comp9 results

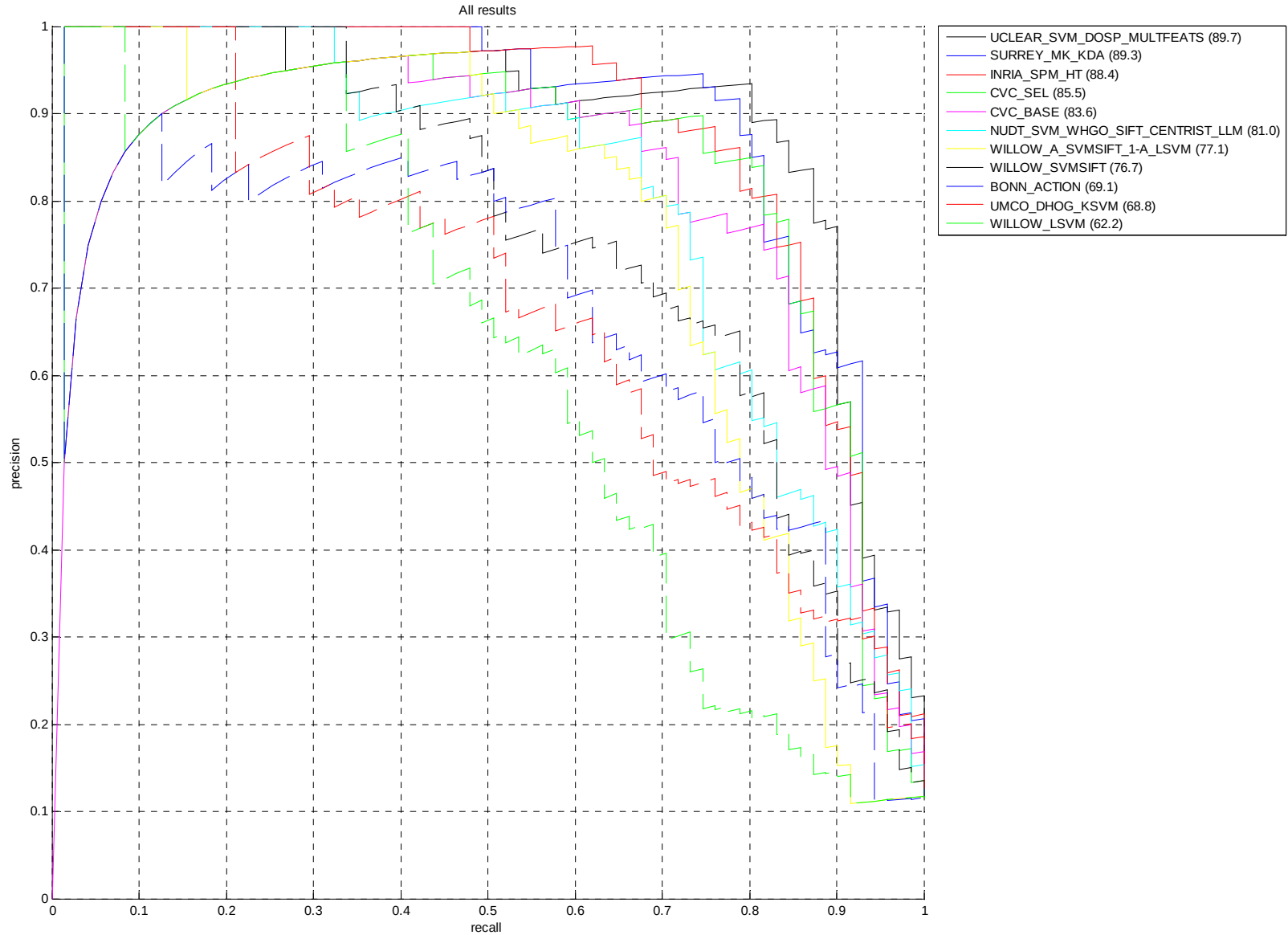
	phoning	playing instrument	reading	riding bike	riding horse	running	taking photo	using computer	walking
BONN_ACTION	47.5	51.1	31.9	64.5	69.1	78.5	32.4	53.9	61.1
CVC_BASE	56.2	56.5	34.7	75.1	83.6	86.5	25.4	60.0	69.2
CVC_SEL	49.8	52.8	34.3	74.2	85.5	85.1	24.9	64.1	72.5
INRIA_SPM_HT	53.2	53.6	30.2	78.2	88.4	84.6	30.4	60.9	61.8
NUDT_SVM_WHGO_SIFT_CENTRIST_LLM	47.2	47.9	24.5	74.2	81.0	79.5	24.9	58.6	71.5
SURREY_MK_KDA	52.6	53.5	35.9	81.0	89.3	86.5	32.8	59.2	68.6
UCLEAR_SVM_DOSP_MULTFEATS	47.0	57.8	26.9	78.8	89.7	87.3	32.5	60.0	70.1
UMCO_DHOG_KSVM	53.5	43.0	32.0	67.9	68.8	83.0	34.1	45.9	60.4
WILLOW_A_SVMSIFT_1-A_LSVM	49.2	37.7	22.2	73.2	77.1	81.7	24.3	53.7	56.9
WILLOW_LSVM	40.4	29.9	32.2	53.5	62.2	73.6	17.6	45.8	41.5
WILLOW_SVMSIFT	47.9	29.1	21.7	53.5	76.7	78.3	26.0	42.9	56.4

(1st, 2nd, 3rd place)

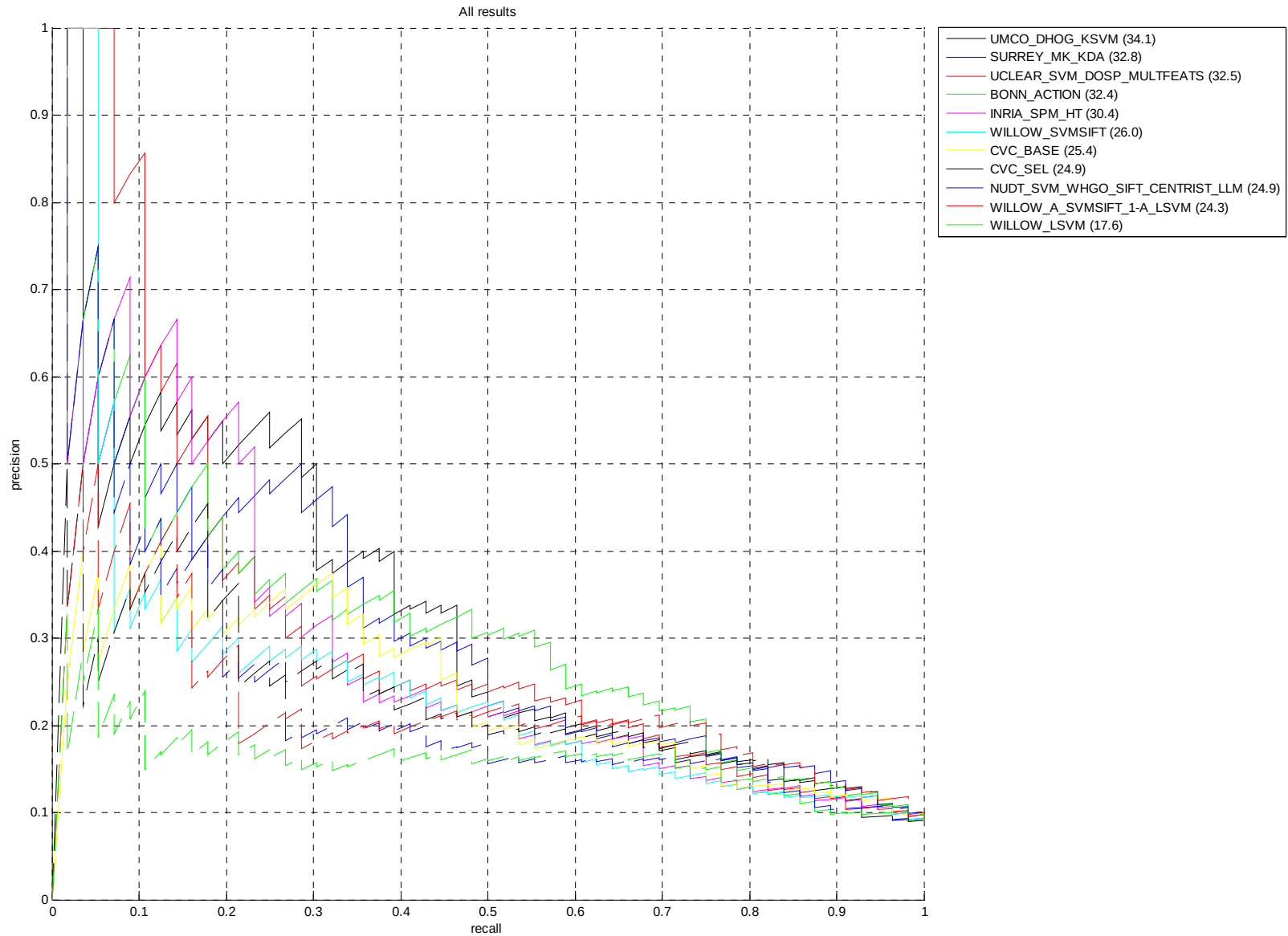
Comp10 results

	phoning	playing instrument	reading	riding bike	riding horse	running	taking photo	using computer	walking
BERKELEY_POSELETS_ACTION	45.9	45.8	23.7	79.9	87.6	83.1	26.2	44.9	66.6

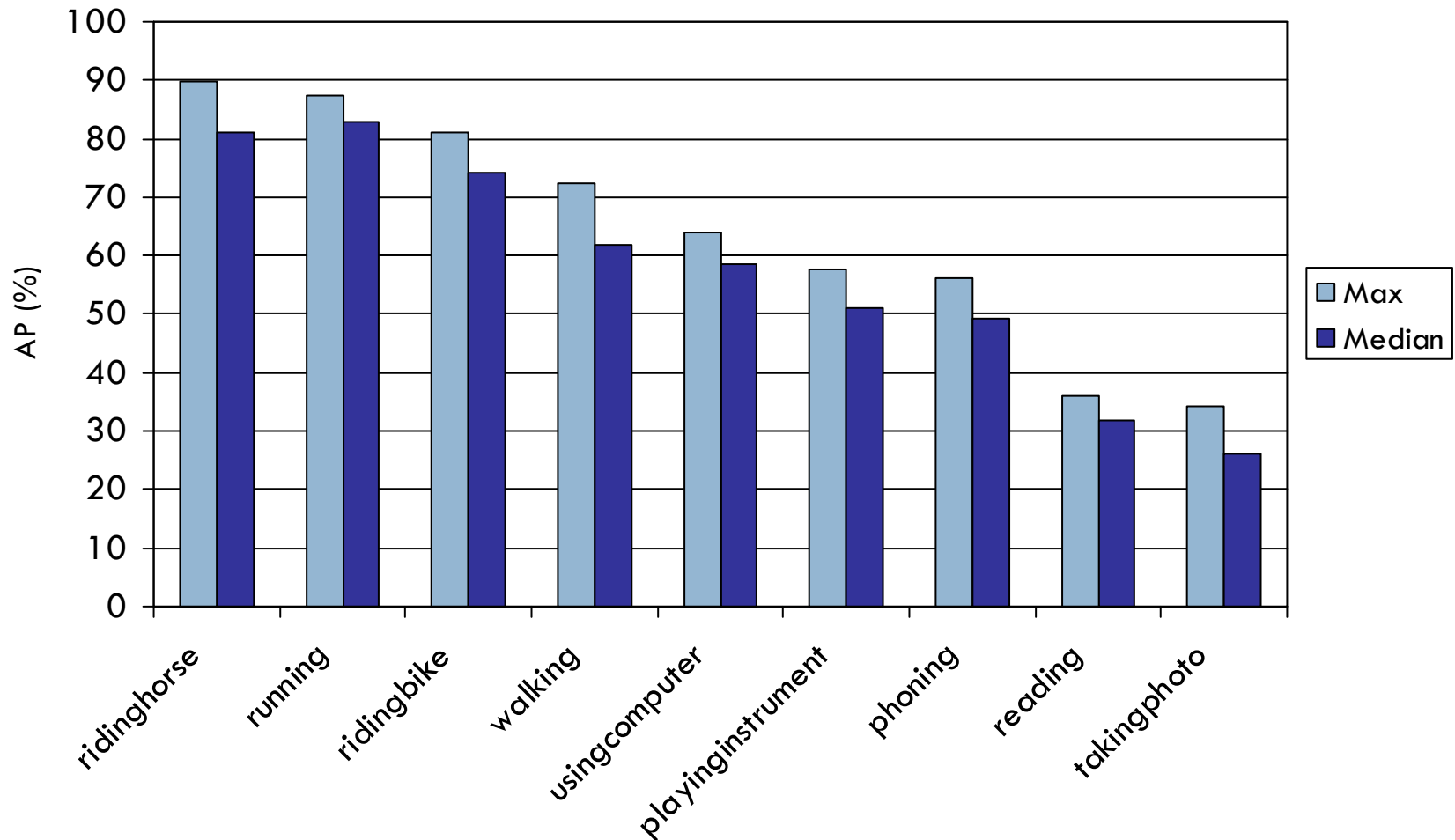
Precision/Recall – Riding Horse



Precision/Recall – Taking Photo

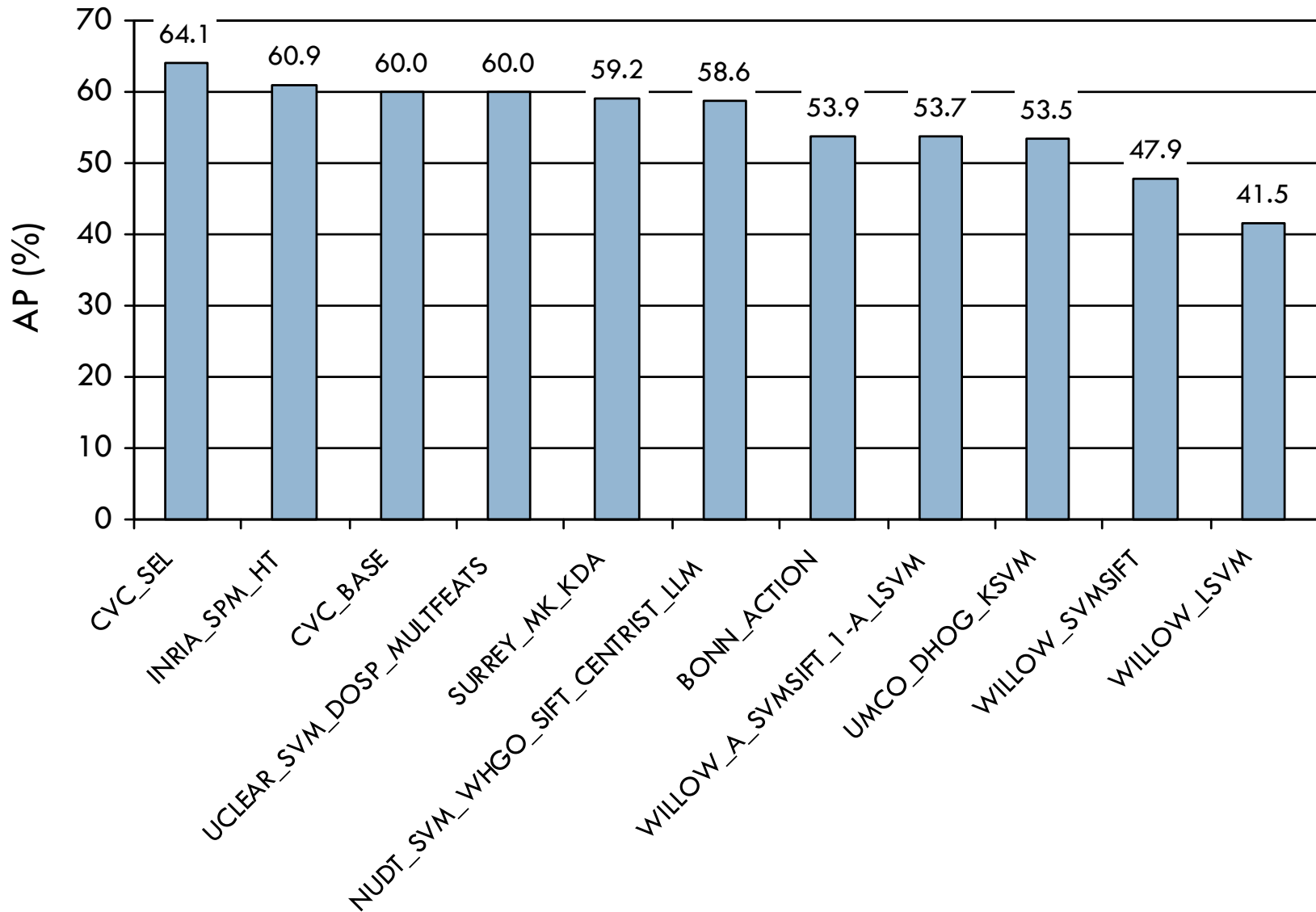


AP by Class



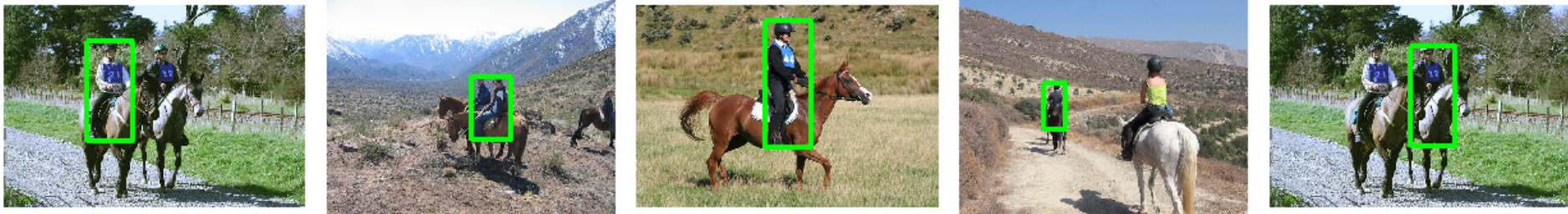
- Max AP: 89.7% (riding horse) ... 34.1% (taking photo)

Median AP by Method



“True Positives”: Riding Horse

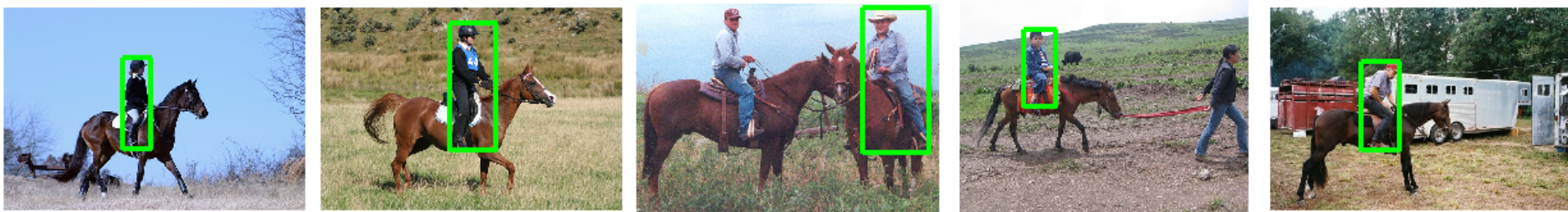
UCLEAR_SVM_DOSP_MULTFEATS



SURREY_MK_KDA

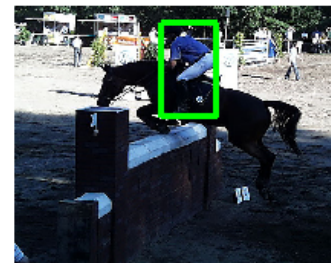
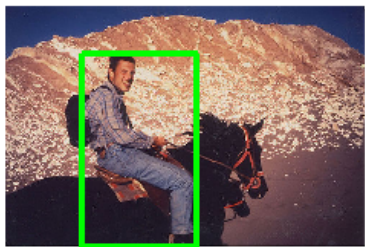
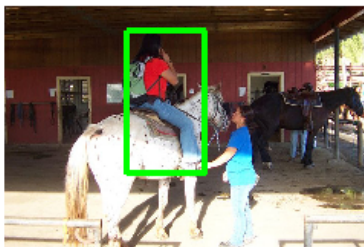


INRIA_SPM_HT

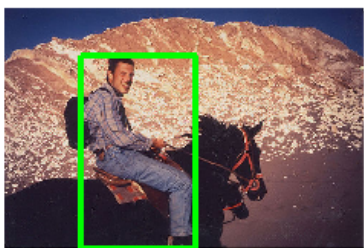


“False Negatives”: Riding Horse

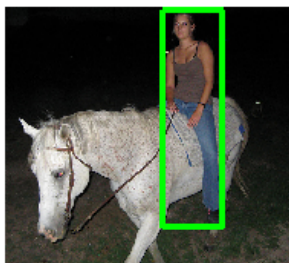
UCLEAR_SVM_DOSP_MULTFEATS



SURREY_MK_KDA



INRIA_SPM_HT

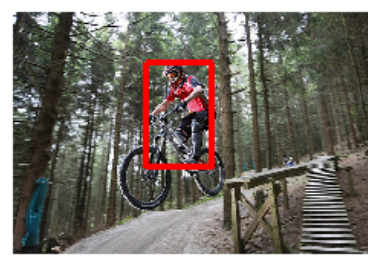


“False Positives”: Riding Horse

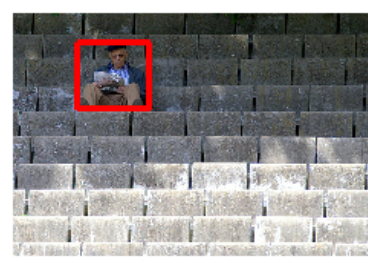
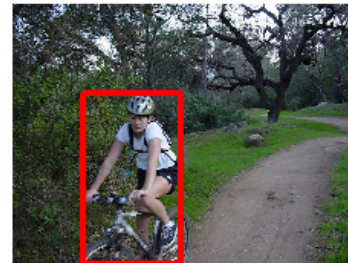
UCLEAR_SVM_DOSP_MULTFEATS



SURREY_MK_KDA



INRIA_SPM_HT

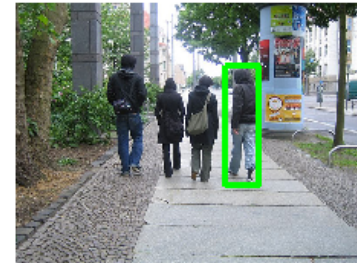


“True Positives”: Walking

CVC_SEL



NUDT_SVM_WHGO_SIFT_CENTRIST_LLM



UCLEAR_SVM_DOSP_MULTFEATS



“False Negatives”: Walking

CVC_SEL



NUDT_SVM_WHGO_SIFT_CENTRIST_LLM

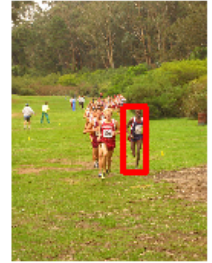
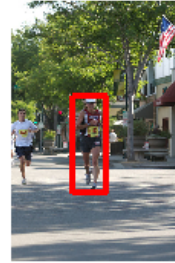


UCLEAR_SVM_DOSP_MULTFEATS



“False Positives”: Walking

CVC_SEL



NUDT_SVM_WHGO_SIFT_CENTRIST_LLM



UCLEAR_SVM_DOSP_MULTFEATS

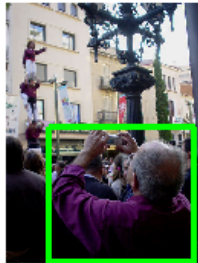


“True Positives”: Taking Photo

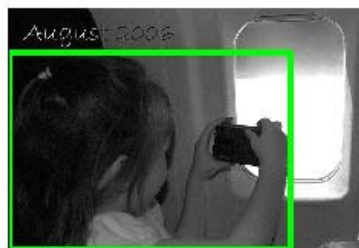
UMCO_DHOG_KSVM



SURREY_MK_KDA



UCLEAR_SVM_DOSP_MULTFEATS



“False Negatives”: Taking Photo

UMCO_DHOG_KSVM



SURREY_MK_KDA



UCLEAR_SVM_DOSP_MULTFEATS



“False Positives”: Taking Photo

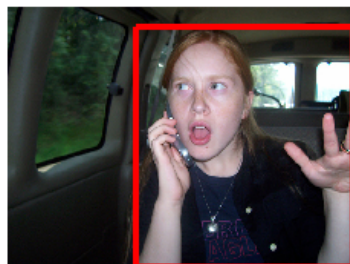
UMCO_DHOG_KSVM



SURREY_MK_KDA



UCLEAR_SVM_DOSP_MULTFEATS



Prizes



■ Joint Winners

■ **CVC_SEL/CVC_BASE**

Nataliya Shapovalova, Wenjuan Gong, Fahad Shahbaz Khan, Josep M. Gonfaus, Marco Pedersoli, Andrew D. Bagdanov, Joost van de Weijer, Jordi González

Computer Vision Center, Universitat Autònoma de Barcelona

■ **INRIA_SPM_HT**

Norberto Adrián Goussies, Arnau Ramisa, Cordelia Schmid
INRIA LEAR

■ **NUDT_SVM_WHGO_SIFT_CENTRIST_LLM**

Li Zhou, Zongtan Zhou, Dewen Hu
National University of Defence Technology

■ **SURREY_MK_KDA**

Piotr Koniusz, Muhammad Atif Tahir, Mark Barnard, Fei Yan, Krystian Mikolajczyk
University of Surrey

■ **UCLEAR_SVM_DOSP_MULTFEATS**

Gaurav Sharma, Frederic Jurie, Cordelia Schmid
University of Caen; INRIA LEAR

■ **UMCO_DHOG_KSVM**

Xutao Lv, Xiaoyu Wang, Xi Zhou, Tony X. Han
University of Missouri - Columbia