Progress with players action recognition

Teo de Campos

CVSSP – Centre for Vision Speach and Signal Processing
Univerisity of Surrey

2^{nd} ACASVA meeting, 13 January 2010

Outline

Outline

Outline

Outline

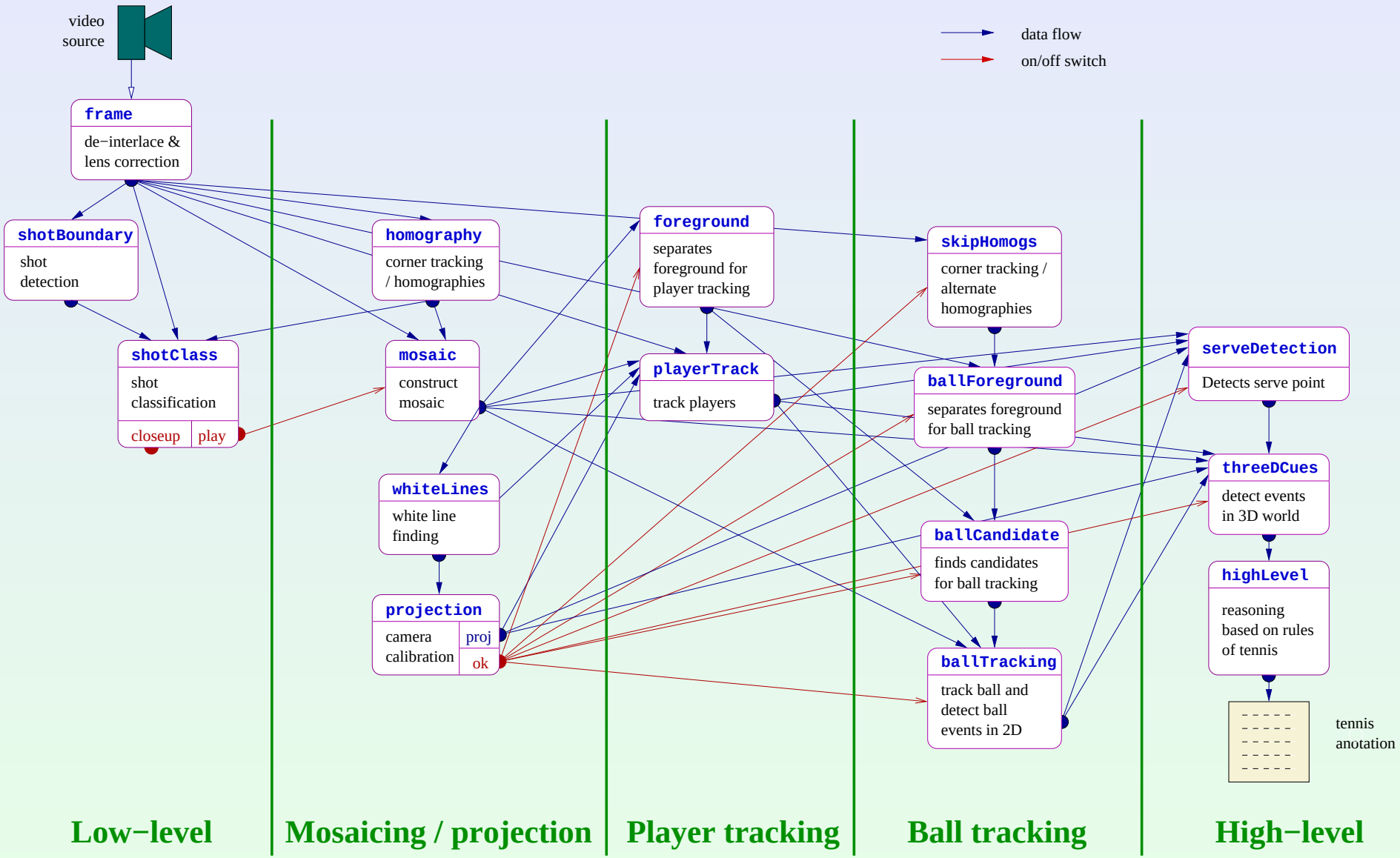
Outline

Tasks for incongruence detection

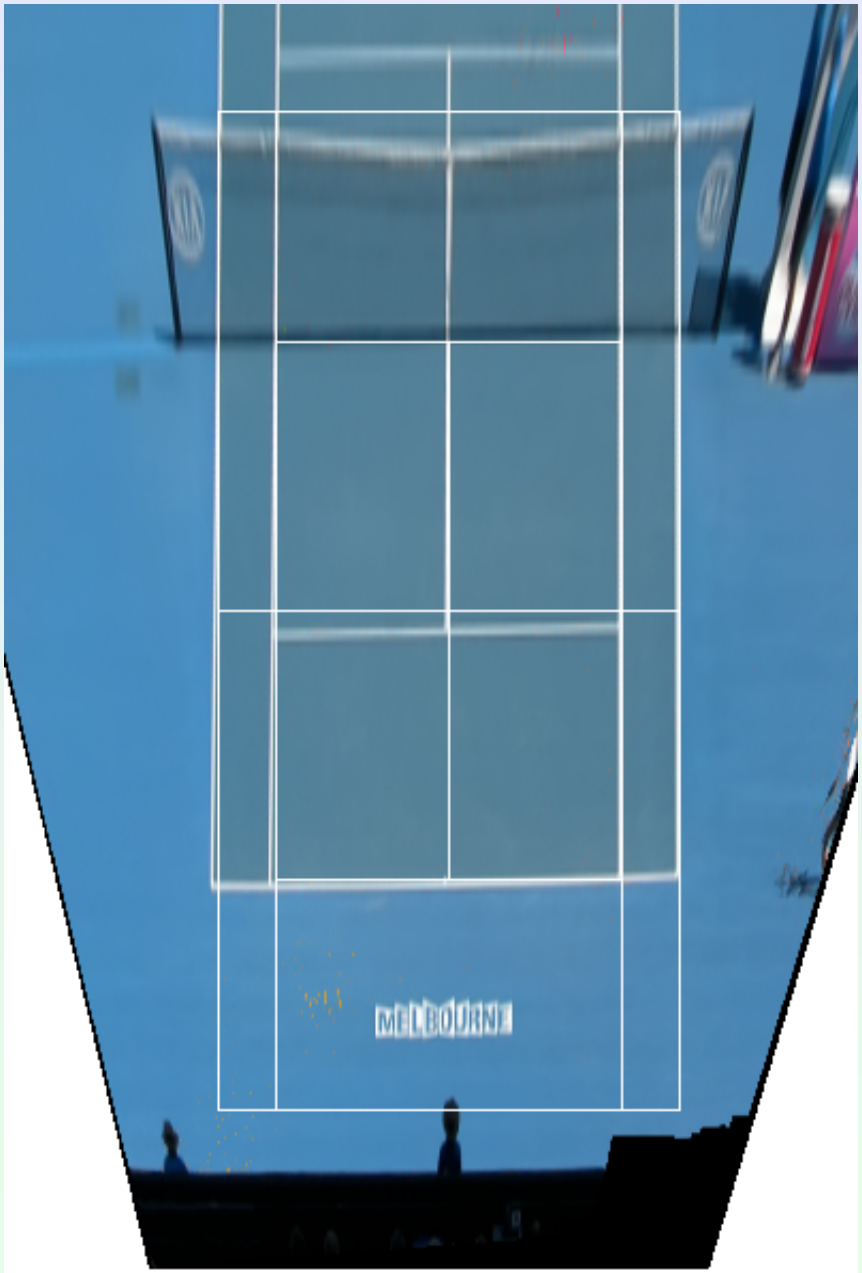
- Prior to rule adaptation, we need to detect incongruent events (Josef's talk).
- For that, we need "weak" classifiers for
  - ball events (Ibrahim)
  - player actions (Teo)
  - audio events
- To train and evaluate the above modules, we need labelled data.
- The VAMPIRE system is a very useful start point.

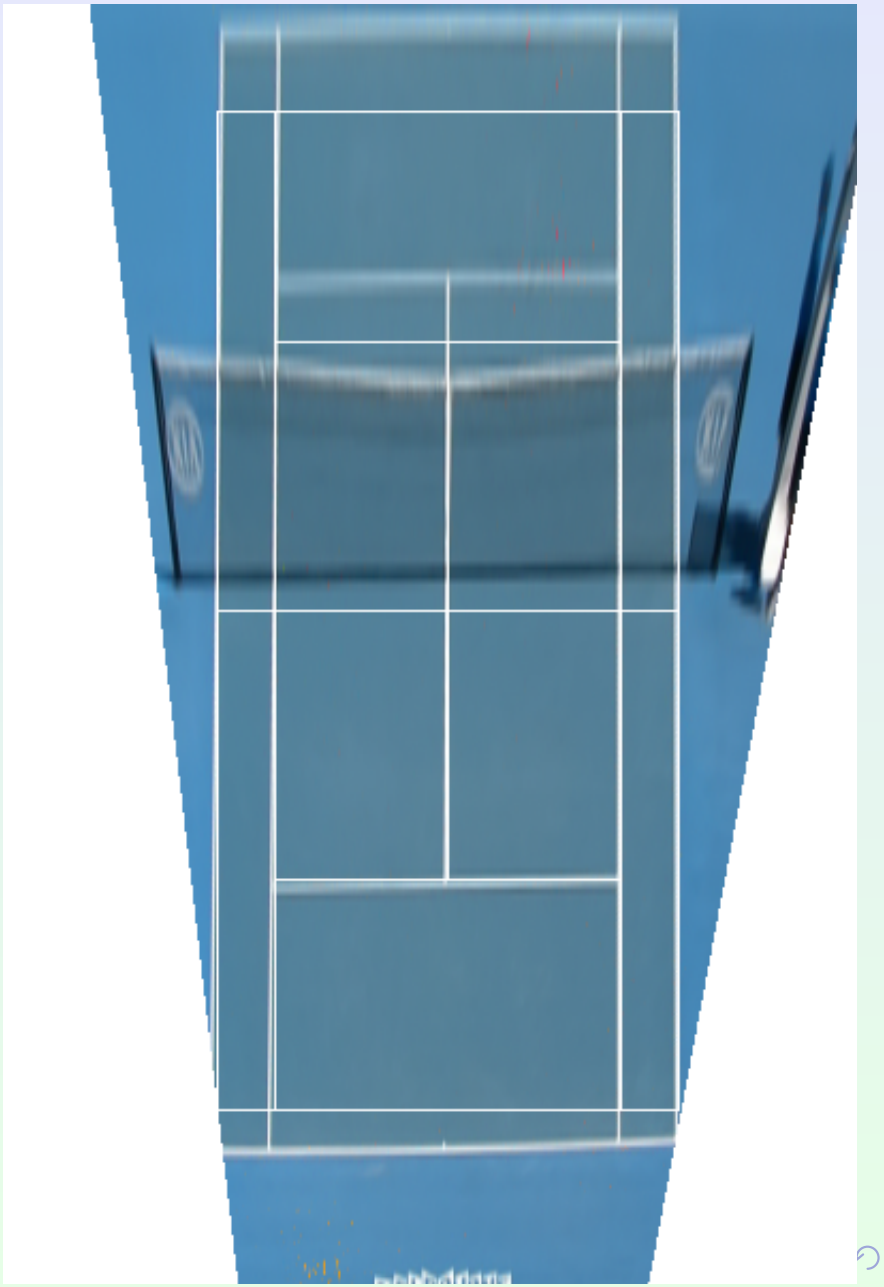
Outline

The VAMPIRE modules

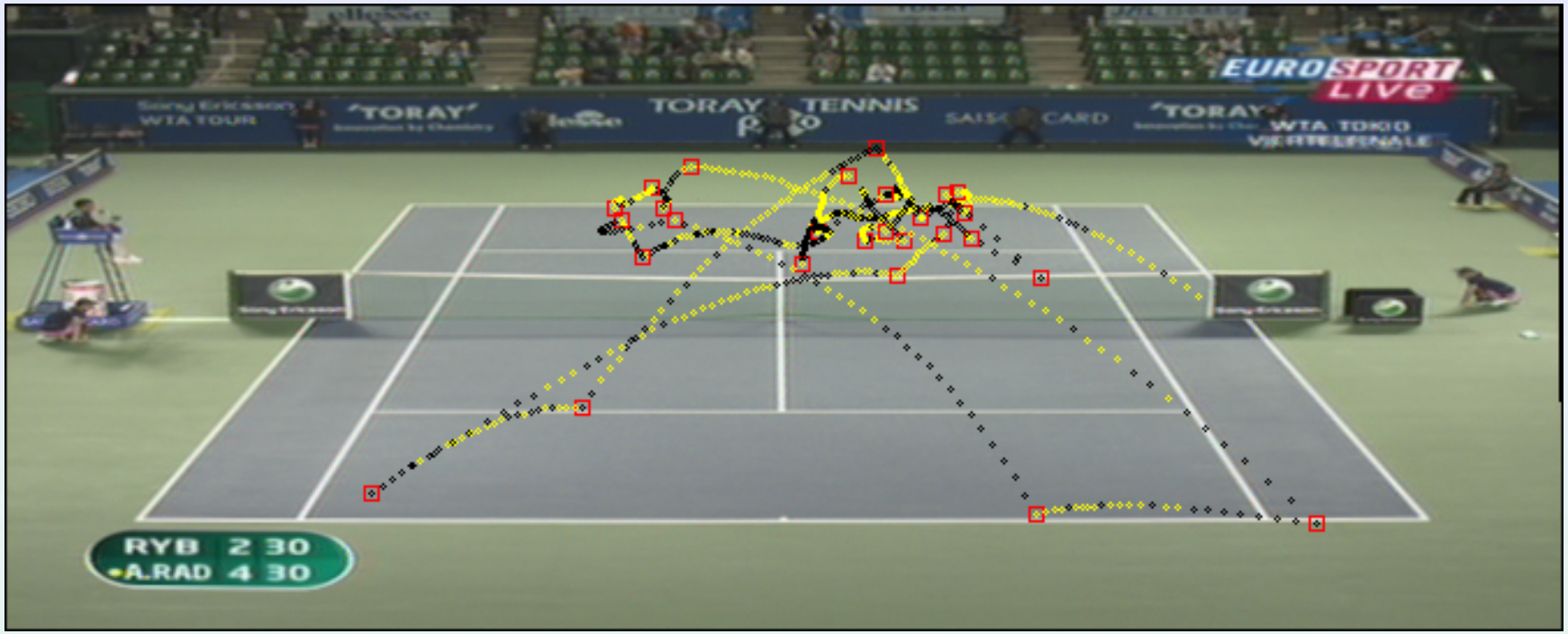


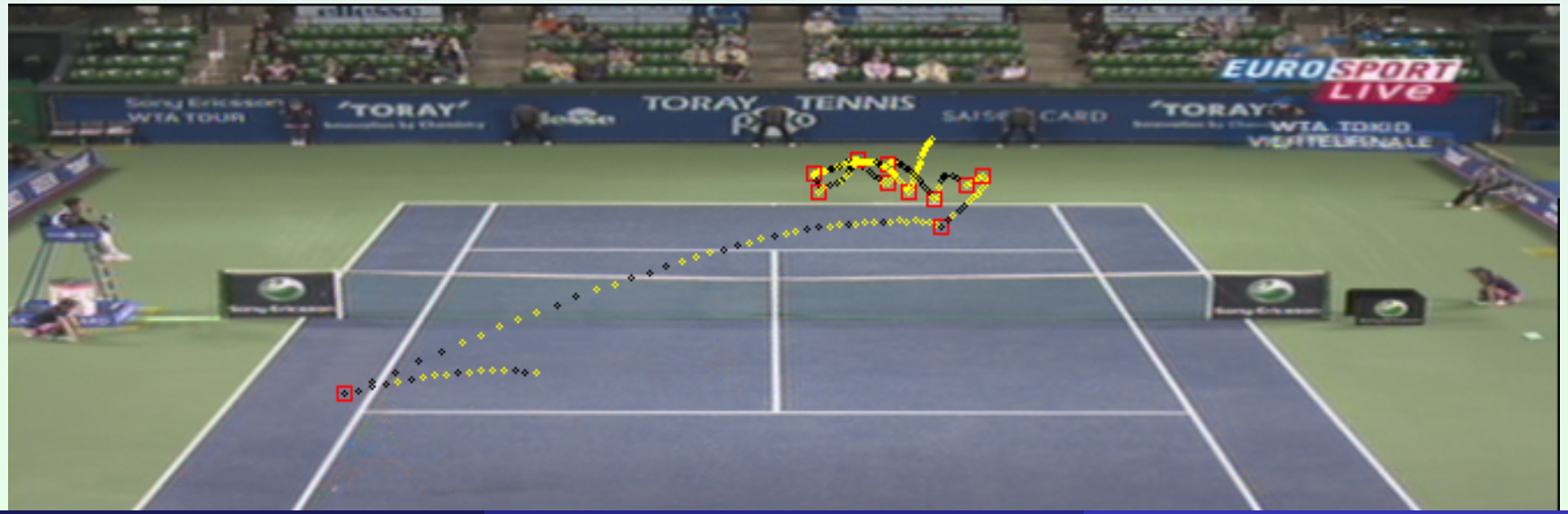
Projection





Ball tracker





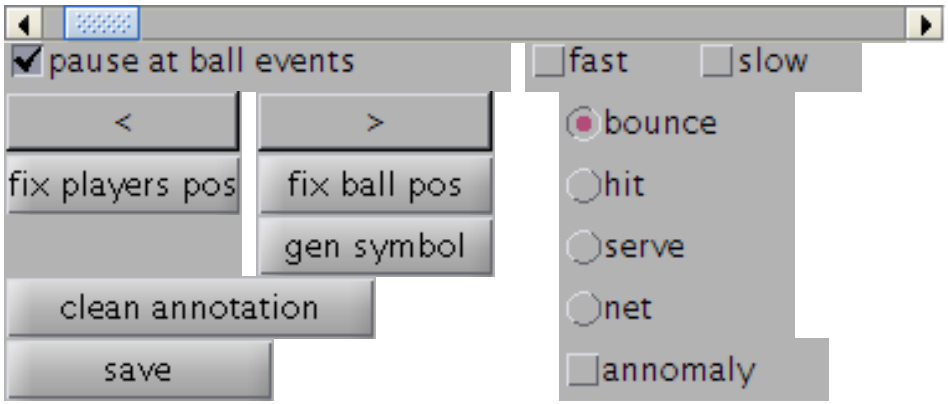
Difficulties with the VAMPIRE system

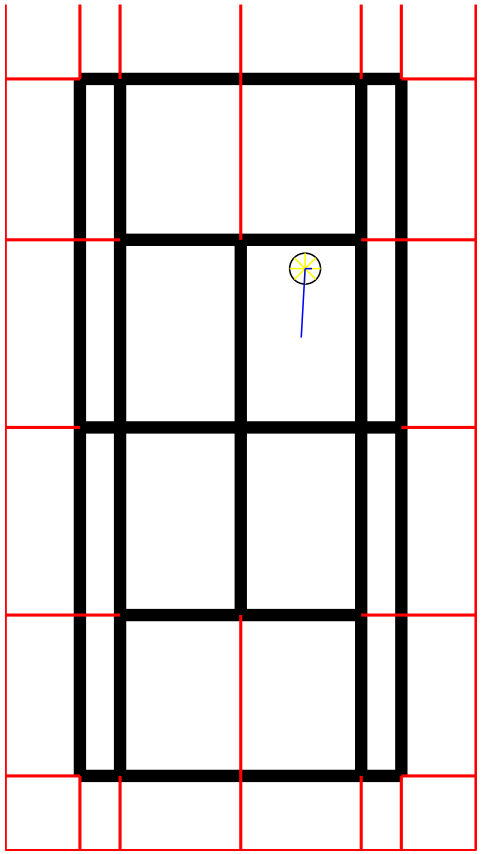
- Shot classification: shots are classified as "play" if $\alpha \cdot \mathit{motion} + \beta \cdot \mathit{hist\_peaks} + \gamma > 0$. Too many shots were taken as "not play". Tuning γ solved this.
- Projection: mismatches between detected lines and court model lines happen too often in doubles. We (mostly Bill) improved by using cropped mosaics, but further work is needed.
- Player tracker: hard coded for games of singles.
- Ball tracker: often outputs too many trajectory change events
- Serve detection: not robust
- High level: hard coded for singles.

Outline

Tennis annotation tool

1482





Data acquisition

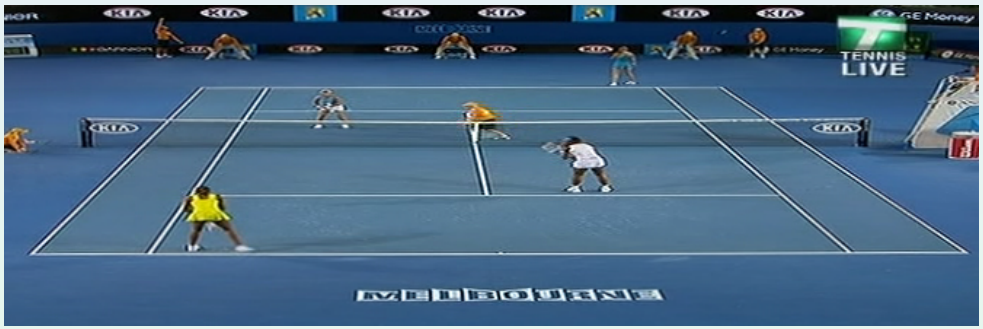
Status	Type	Year	Players	Scores	Lang	Tournament
Run	Mens Sngl	2003				Aus. open (TV)
Done	Ladies Sngl	2003	Williams sisters			Aus. open (TV)
Done	Ladies Sngl	2009				Japan (TV)
	Ladies Sngl	2009				(TV)
Done	Ladies Dbl	2008				Australian open
Done	Ladies Dbl	2009	Williams sisters vs.			Australian open
	Mens Dbl	2008	Nestor/Zimonjic Def. Bjorkman/Ullyett	76 67 63 63	Eng	Wimbledon
	Ladies Dbl	2009	S.Williams/V.Williams Def. Stosur/Stubbs	76 64	Eng	Wimbledon
	Ladies Dbl	2007	Black/Huber Def. Srebotnik/Sugiyama	57 63 108 (TB)	Eng	Masters
	Ladies Dbl	2002	Hingis/Kournikova Def. Hantuchova/Sanchez	62 67 61	Eng	Australian open
	Mens Dbl	2009	B.Bryan/M.Bryan Def. Bhupathi/Knowles	64 64	Eng	Masters
	Mens Dbl	2009	Lopez/Verdasco Def. Berdych/Stepanek	76 75 62	Eng	Davis and Fed Cup
	Mens Dbl	2008	B.Bryan/M.Bryan Def. Dlouhy/Paes	76 76	Eng	US Open
	Mens Dbl	2009	Dlouhy/Paes Def. Moodie/Norman	36 63 62	Dutch	Roland Garros
	Ladies Dbl	1999	S.Williams/V.Williams Def. Hingis/Kournikova	63 67 86	Jap	Roland Garros
	Ladies Dbl	2007	Dechy / Safina Def. Chan / Chuang	64 62	Fr	US Open

- We grabbed four games and purchased DVDs of 12 games.
- Disk space has become an issue! We need about 5TB to just to pre-process (with VAMPIRE) all of the above.

Outline

Processing foreground blobs for player detection

-



=

Mosaic, built per shot

-

Input image: de-interlaced field with radial distortion corrected, registered with the mosaic

=

Result: moving blobs. These are merged with a morphological opening operation

Processing foreground blobs for player detection

Algorithm

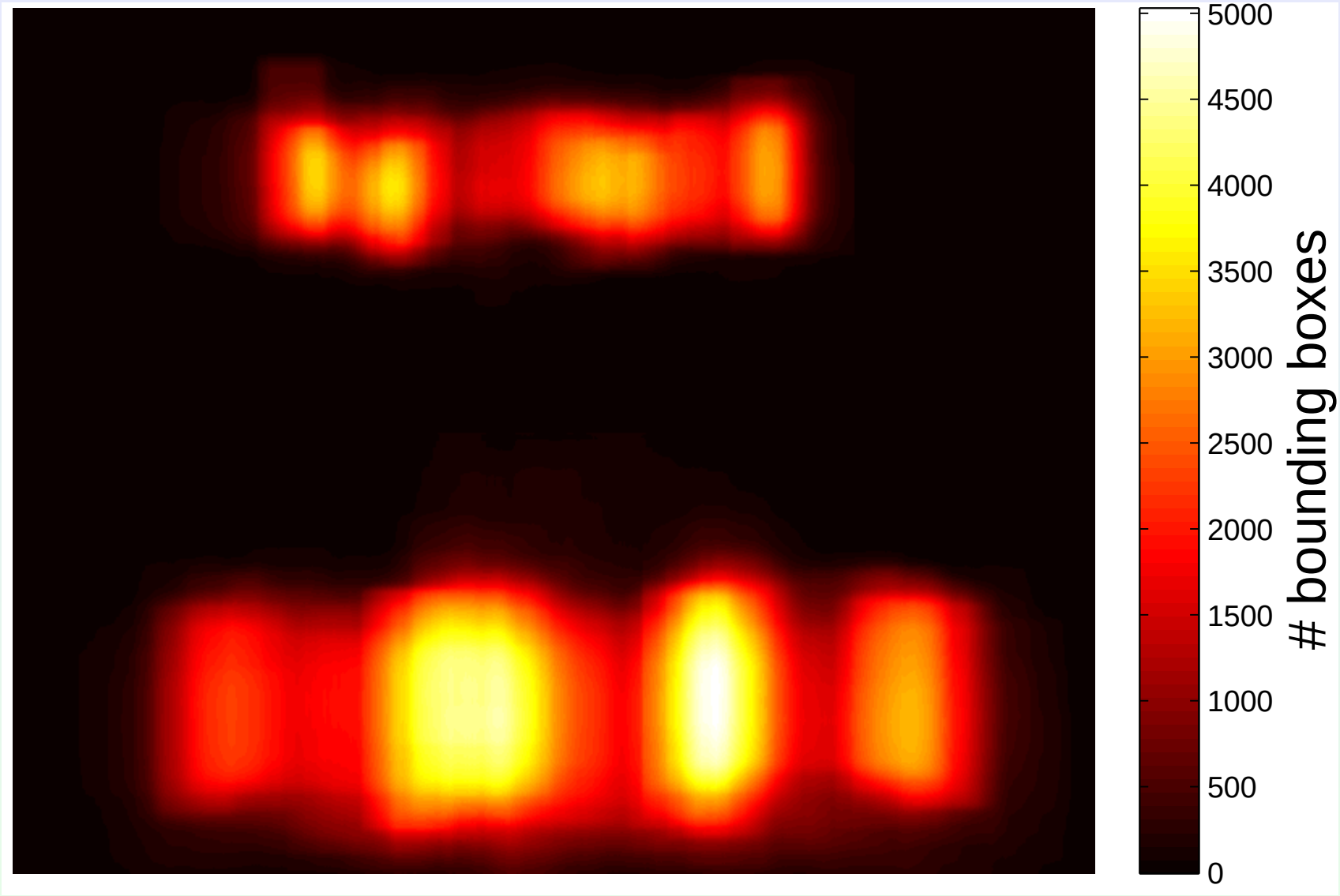
1. Background subtraction
2. Morphological opening
3. Fit bounding boxes to all continuous blobs
4. Merge nearby boxes
5. Apply geometric constraints: area, aspect ratio, ratio area/BB_area
6. Apply temporal constraint
7. Apply foreground mask

Processing foreground blobs for player detection

Resulting boxes for the previous image

- Initial background subtraction and pre-processing: 119 red boxes
- Merging: 32 cyan boxes
- Geometric constraints: 8 dashed magenta boxes
- Spatio-temporal consistence: 7 dashed green boxes
- Mask filter: 5 dotted yellow boxes

Player location pdf



Computed from a 35 minutes footage of singles.

Foreground mask

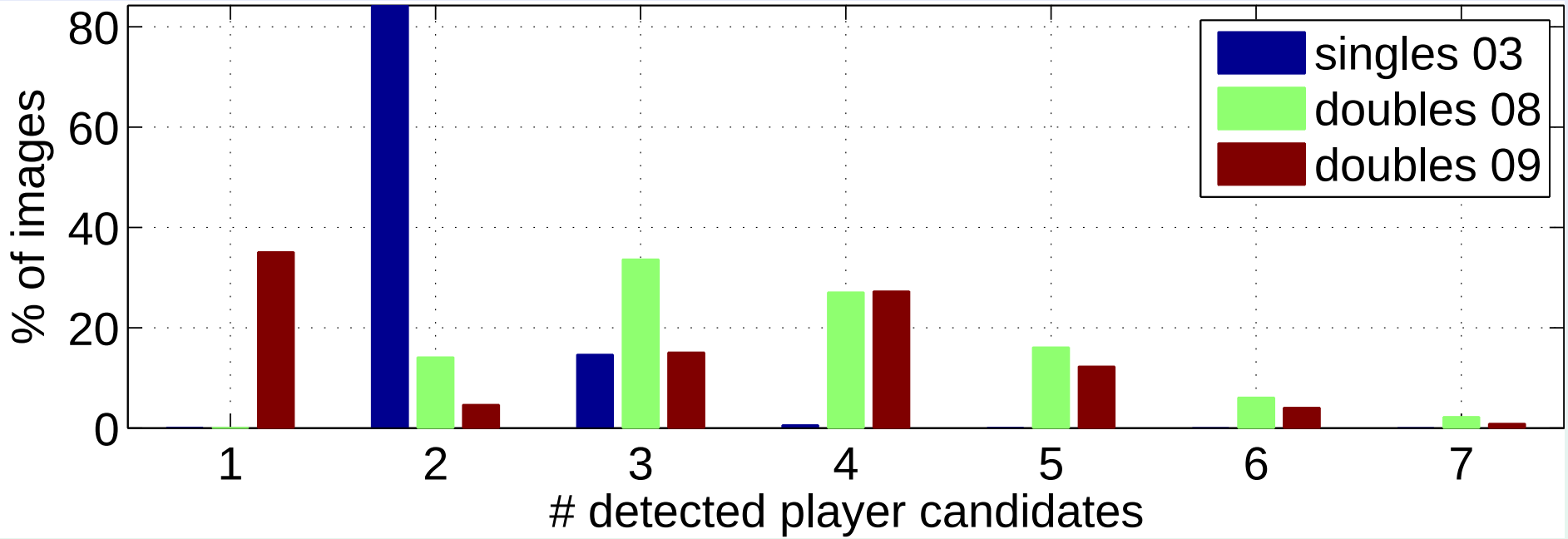
Statistics of the bounding boxes

Footage	BS blobs	motion	mask
singles 03	177.4 ± 23.8	2.8 ± 1.5	2.2 ± 1.3
doubles 08	64.9 ± 21.9	4.7 ± 1.3	3.8 ± 1.2
doubles 09	50.4 ± 44.7	3.5 ± 2.2	3.0 ± 1.7

Number of player candidates per frame after some stages of the processing pipeline.

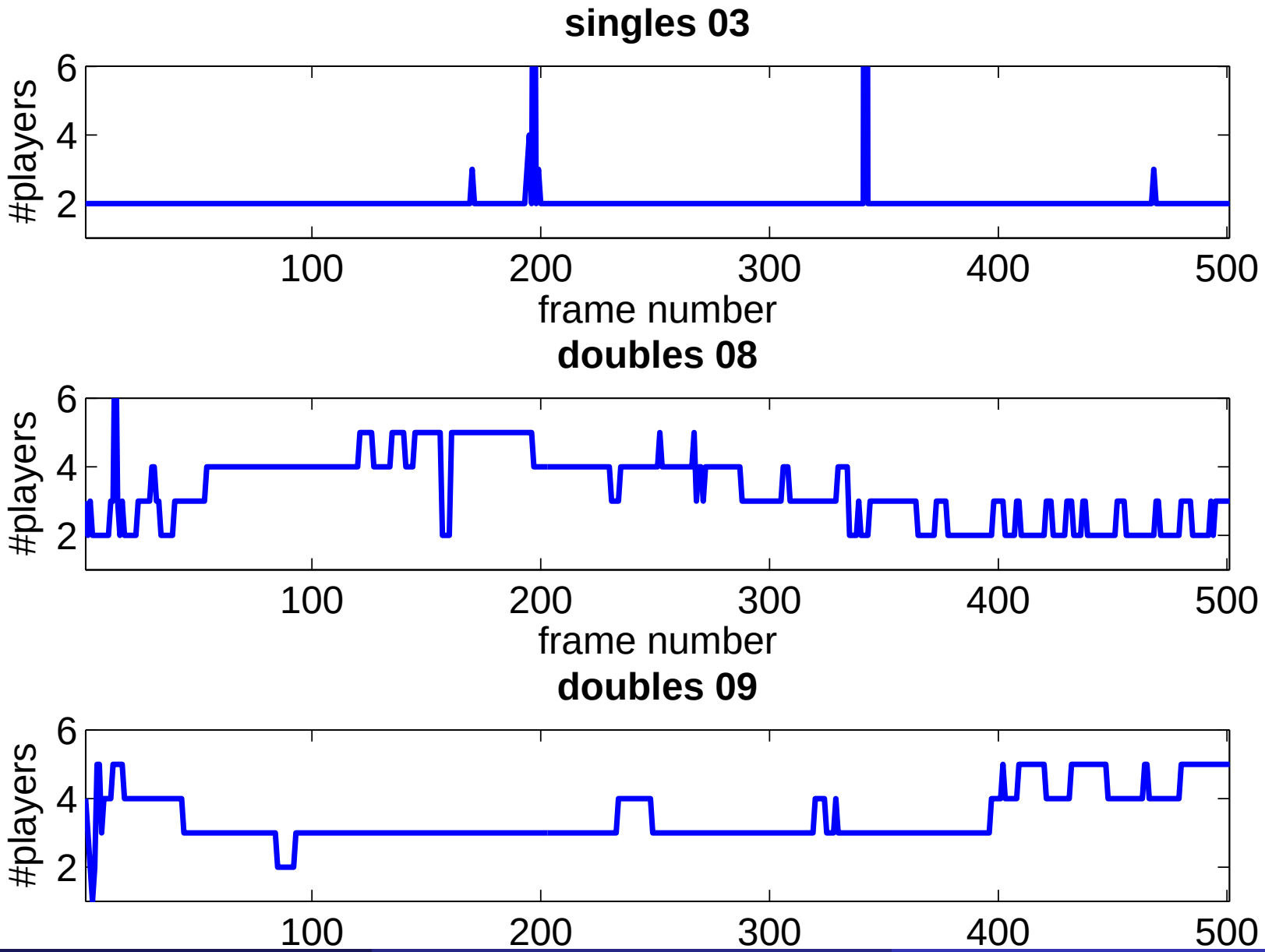
- *BS blobs*: initial blob detection from background subtraction;
- *motion*: application of a motion smoothness constraint;
- *mask*: application of the likely player location mask.

Statistics of the bounding boxes

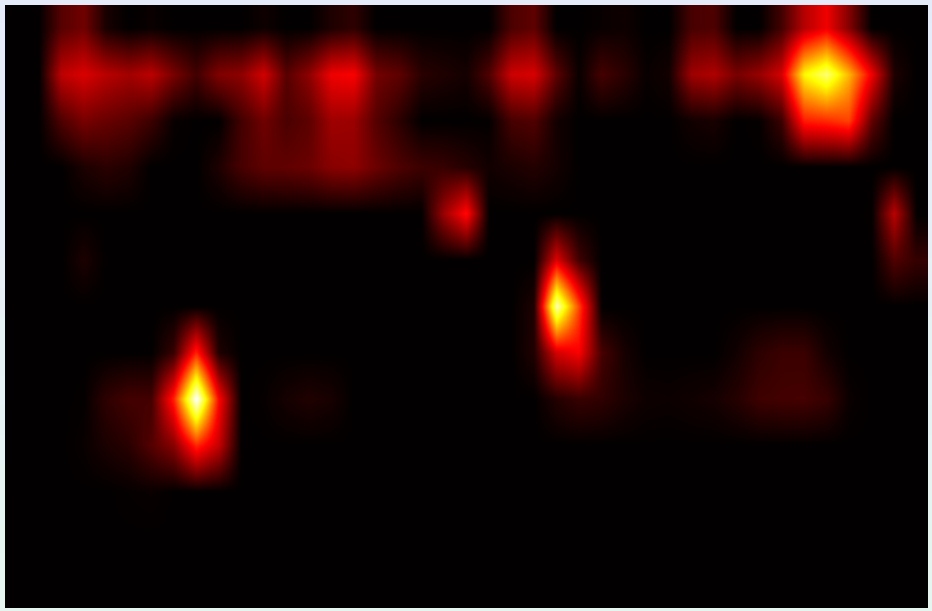


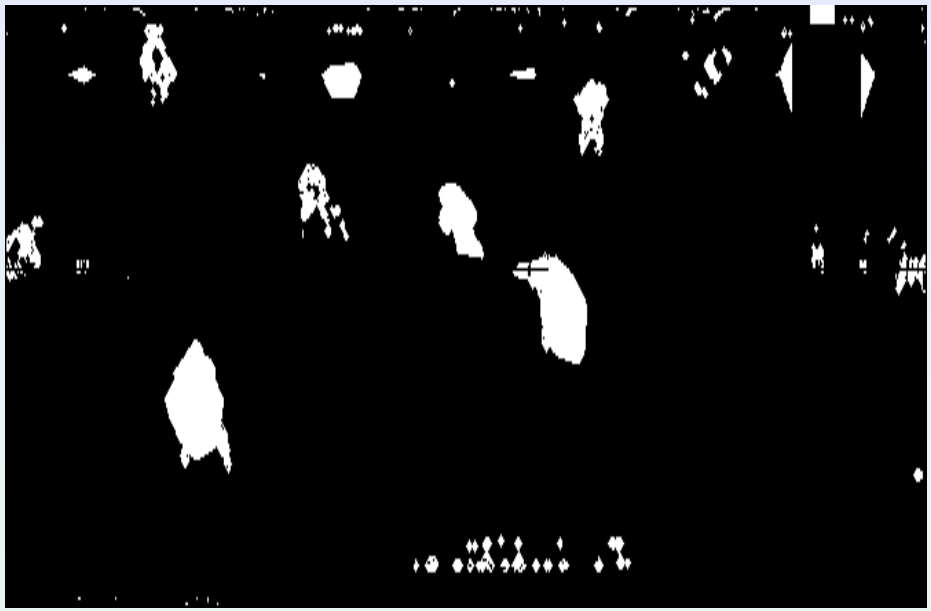
Detected player candidates in each frame of play shots.
Total number of play frames: 25984, 17737 and 33223 for sngl03, dobl08 and dobl09, respectively.

Player detections over time



Combining with visual saliency

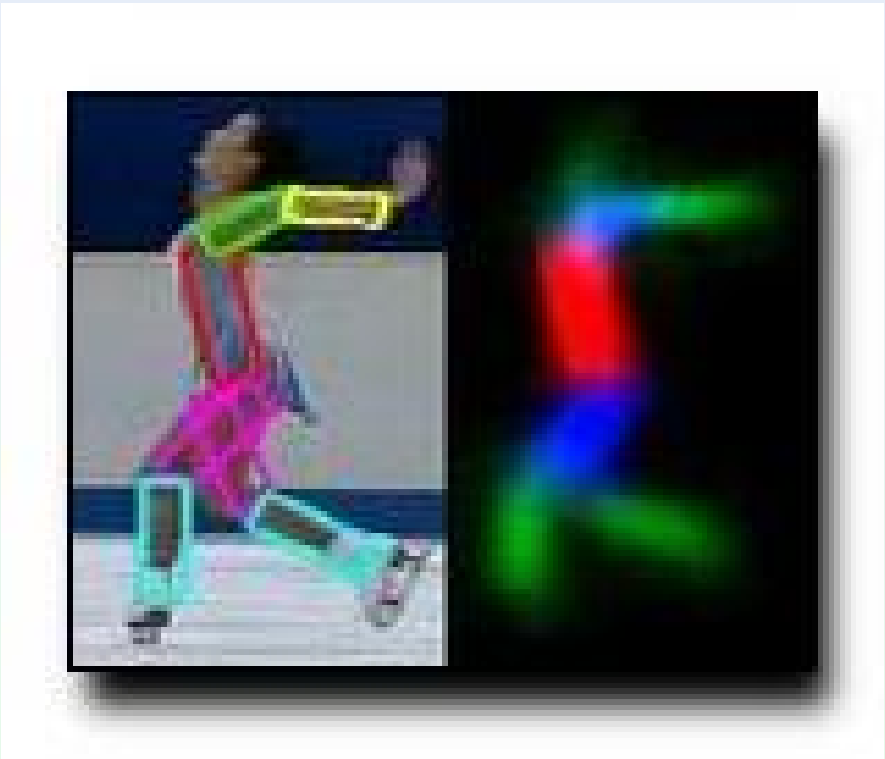




Combining background subtraction with visual saliency maps from [Walther & Koch, 2006] by thresholding both and using an OR operation.

Outline

Using parts-based person detector to locate players [Ramanan *et al.*, 2007]



Results training with a serve frame

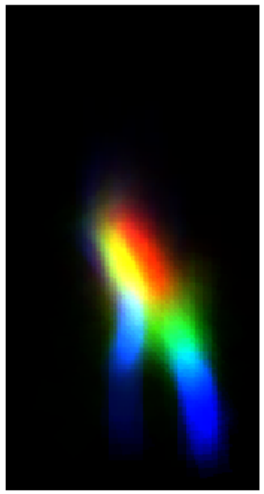
Frame 1

Torso pixels

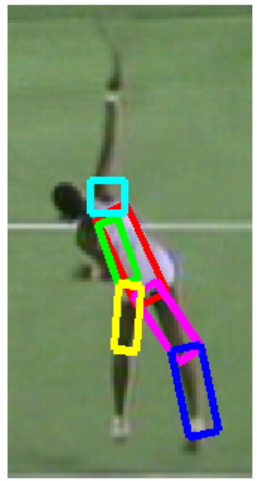
Lower arm pixels

Lower leg pixels

Posterior



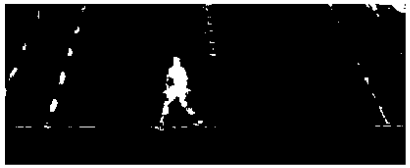
Mode in posterior



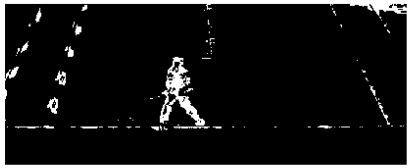
Results training with “walking person”

Frame 9567

Torso pixels

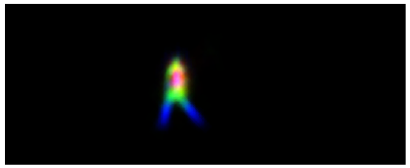


Lower arm pixels

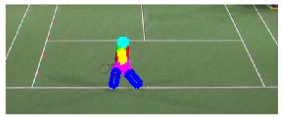


Lower leg pixels

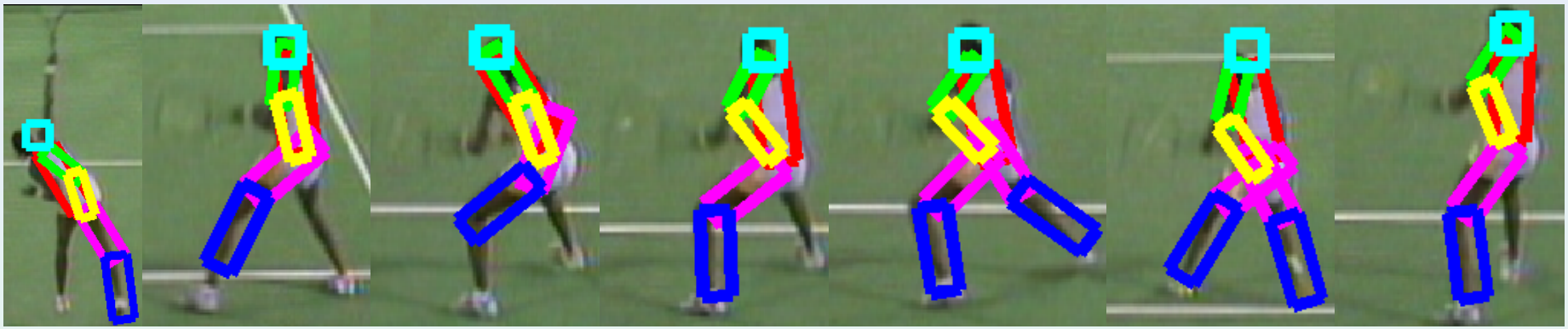
Posterior



Mode in posterior



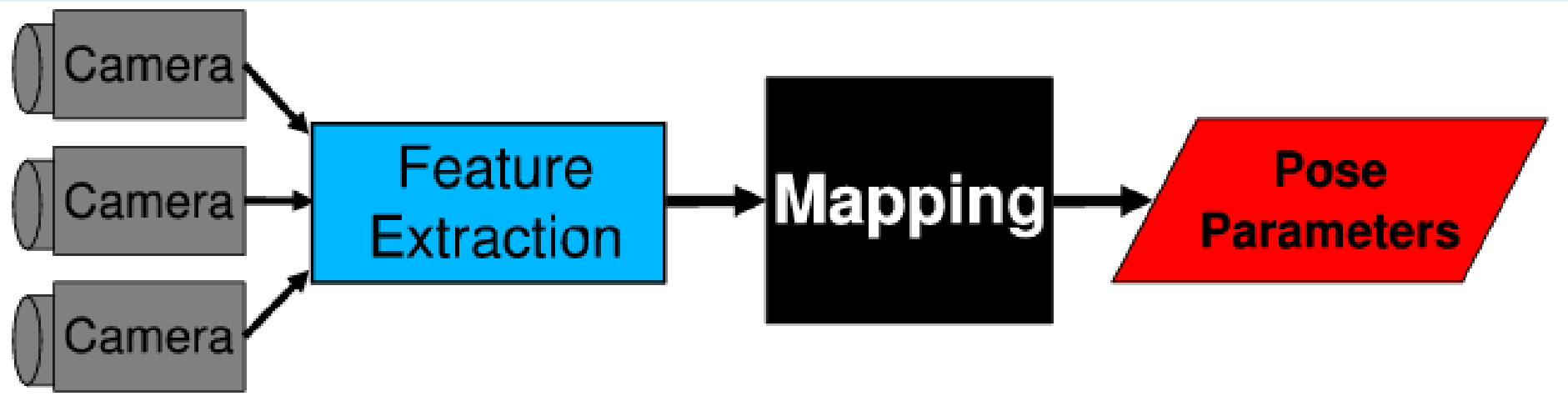
More results



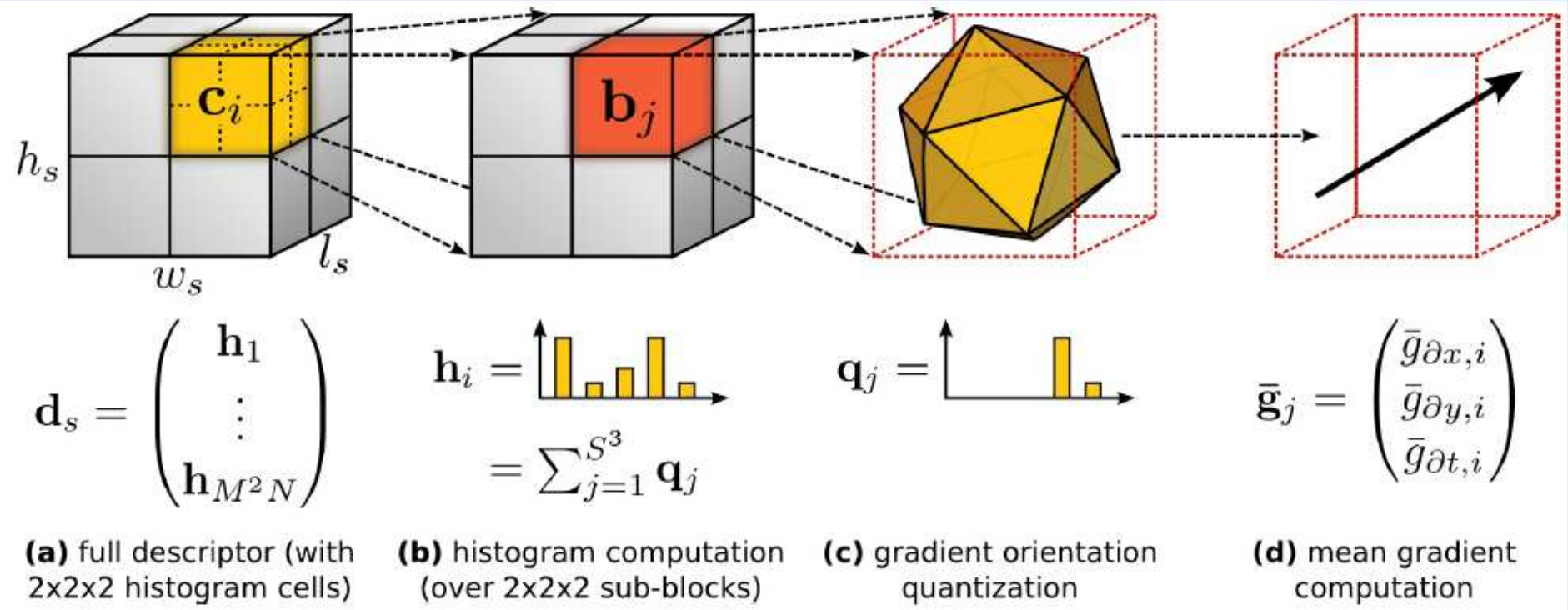
- Bad to locate limbs and articulations
- Good to detect people

Action Recognition

Using the “discriminative” approach.



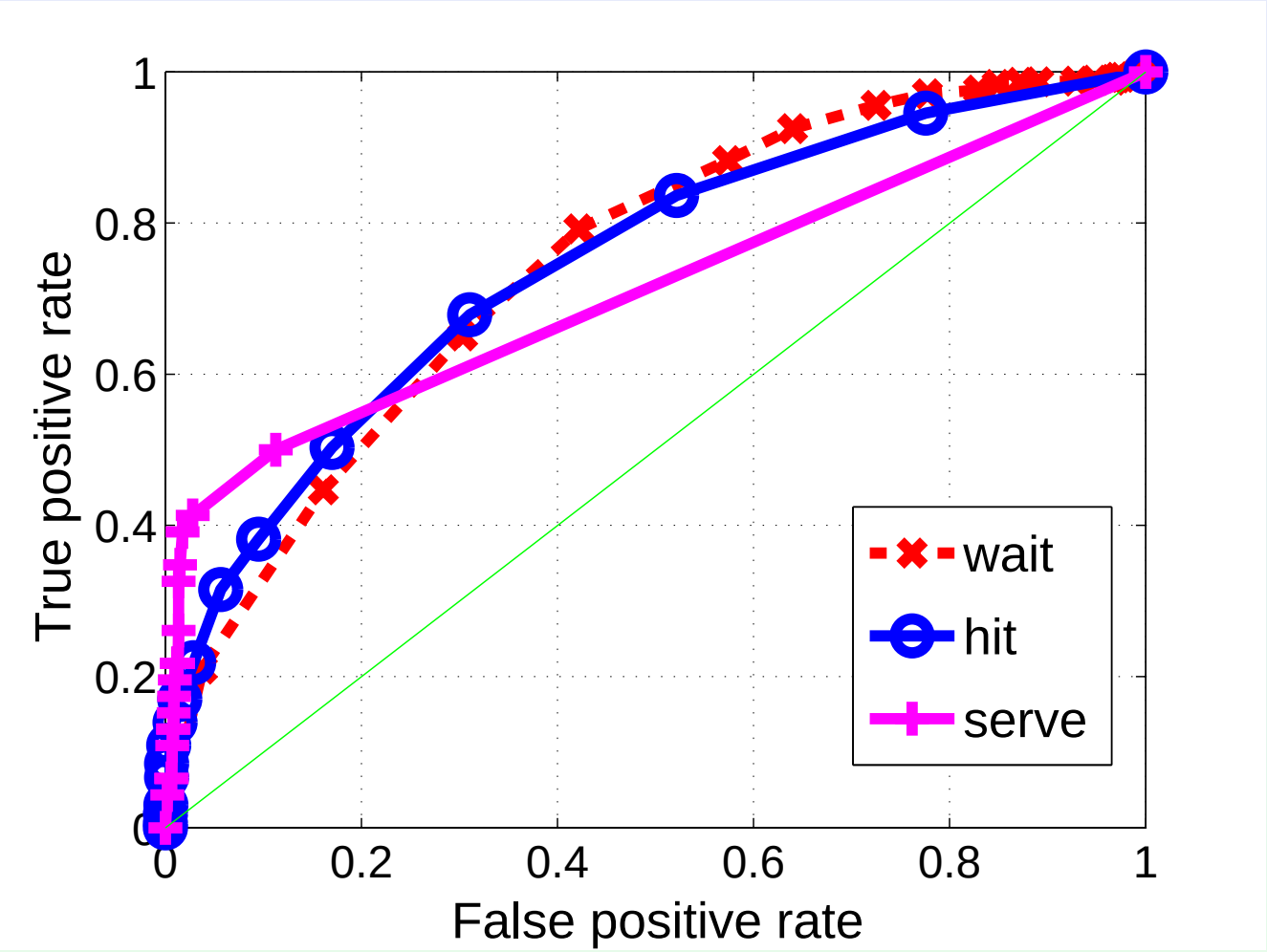
[Kläser *et al.*, 2008]'s 3D SIFT



Proven to be among the state-of-the-art descriptor in [Wang *et al.*, 2009]

Direct classification – applying PCA on the descriptors

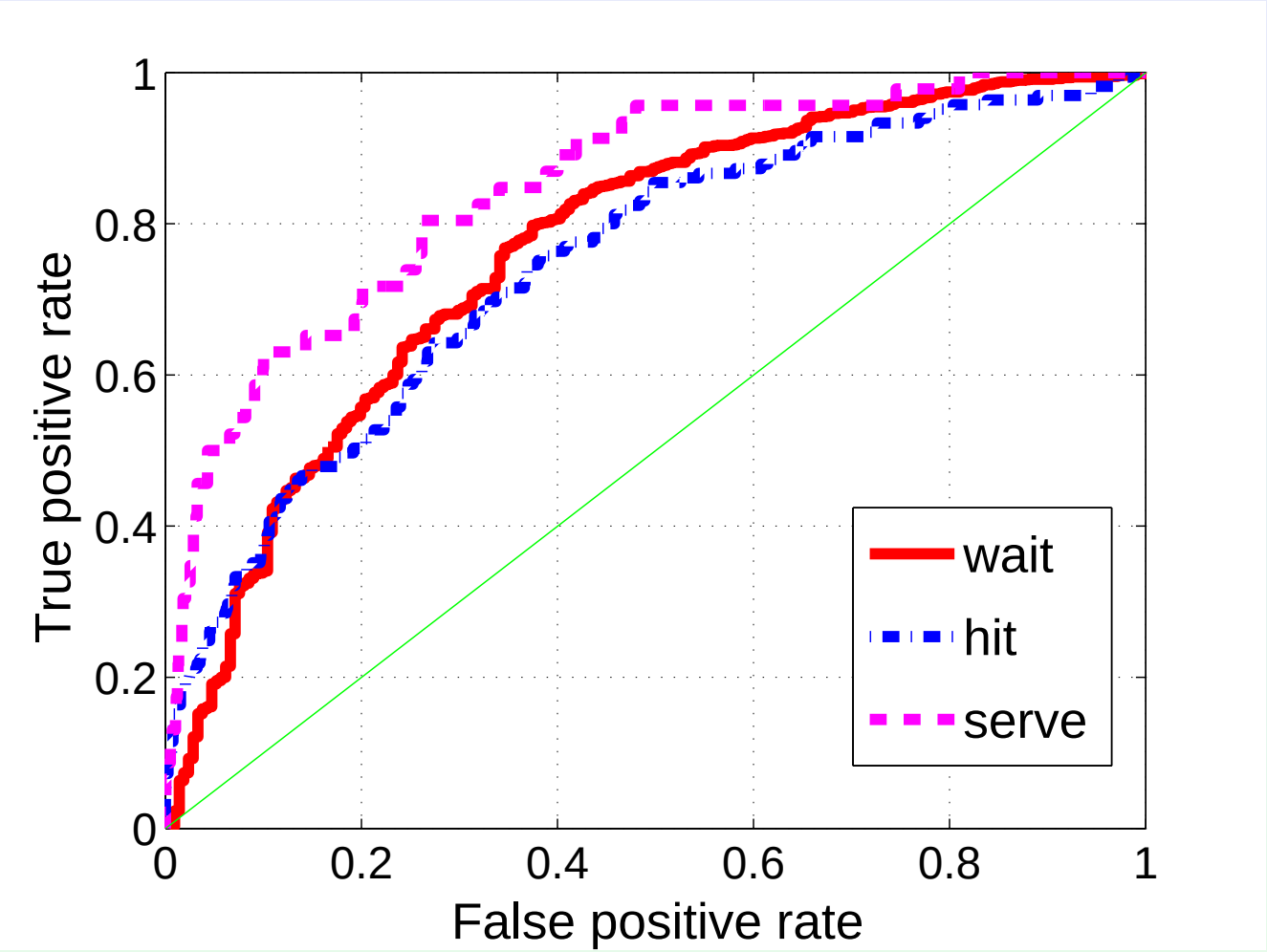
K Nearest Neighbours (K=21)



ROC curves. Mean AUC: 73.40%.

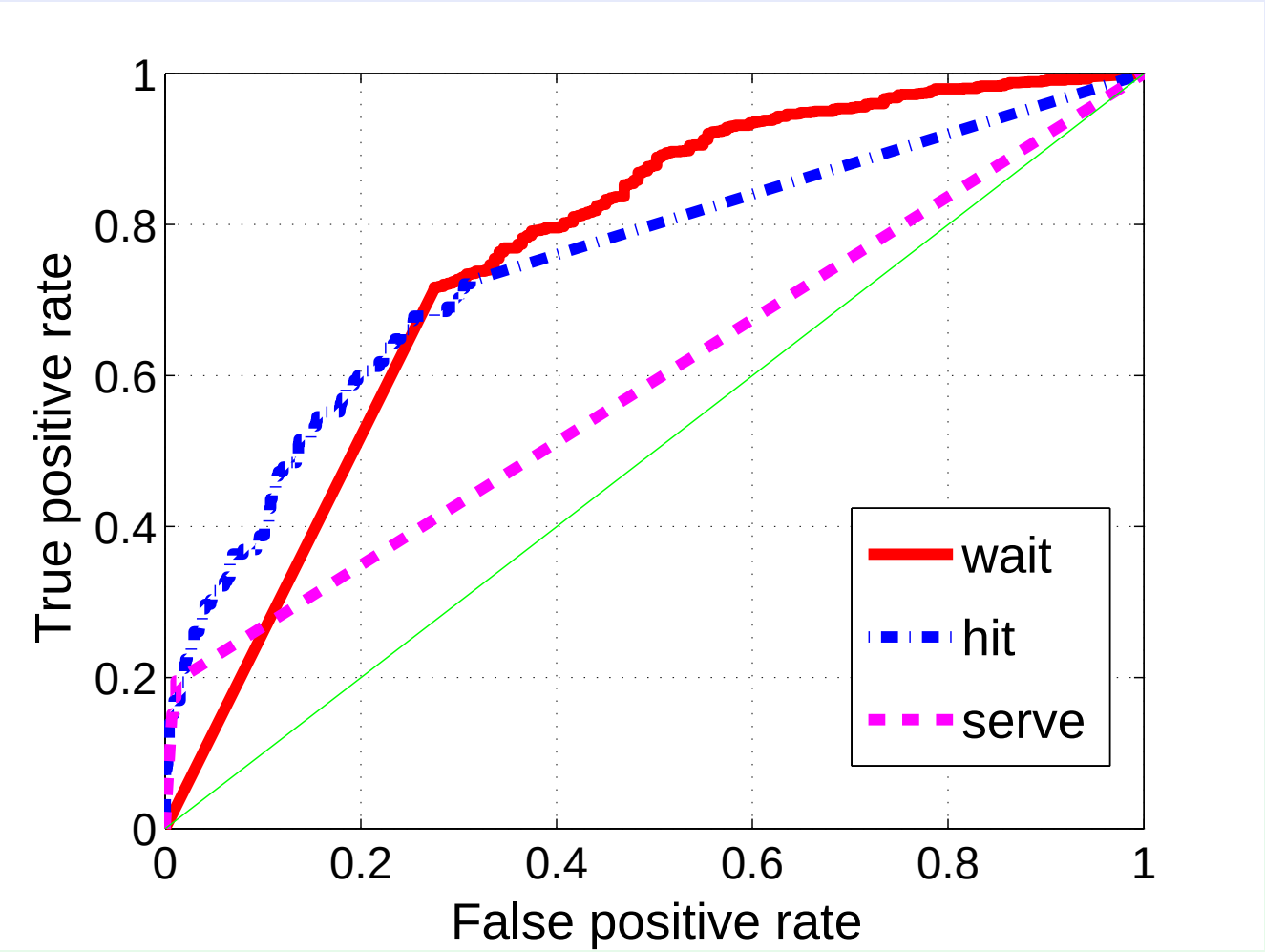
Features: square in space (same height as the player height), 12 frames in time

Action classification – MLP Neural Net



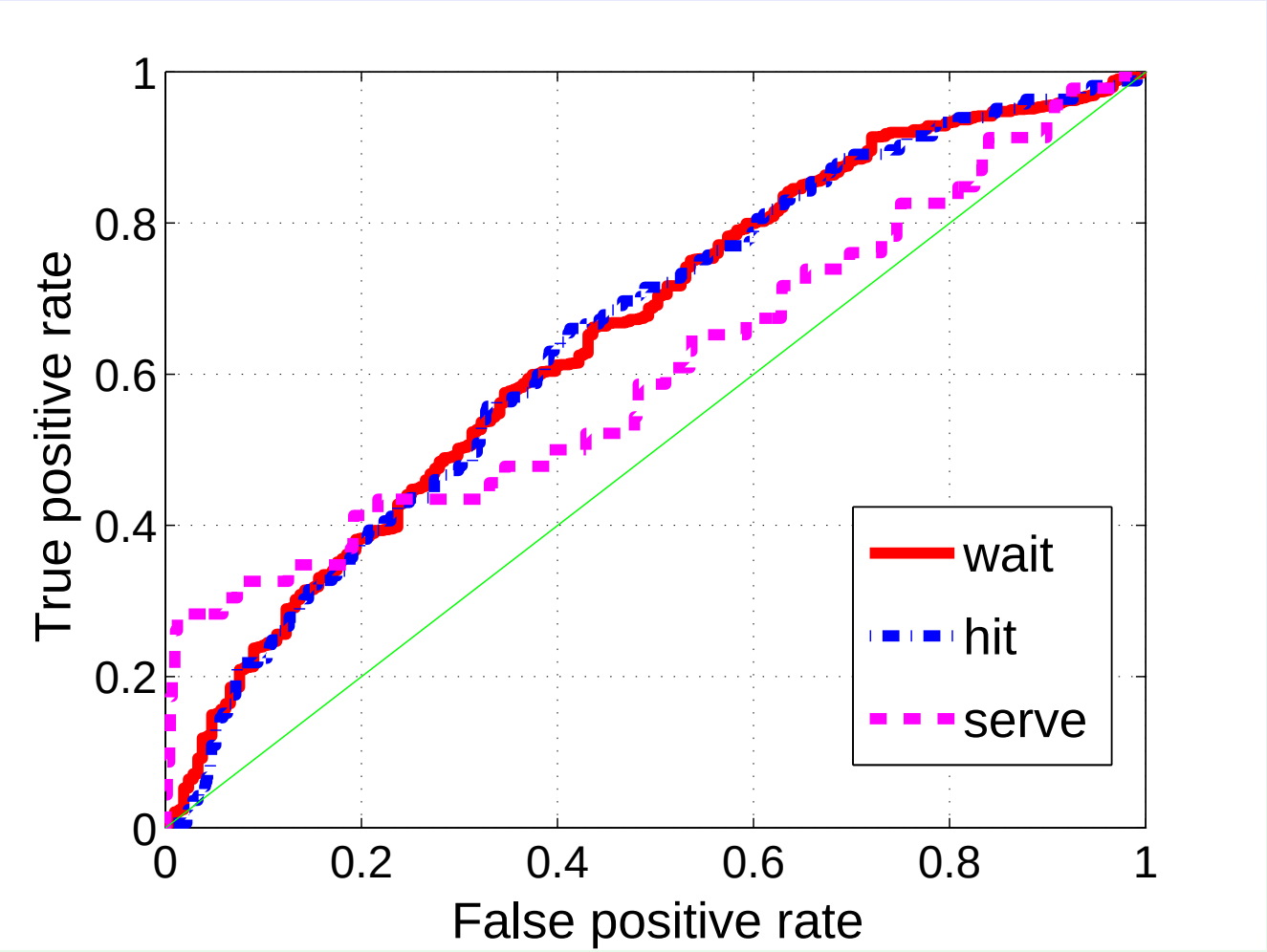
Mean AUC: 78.57%.

Action Classification – SVM linear



Mean AUC: 69.89%.

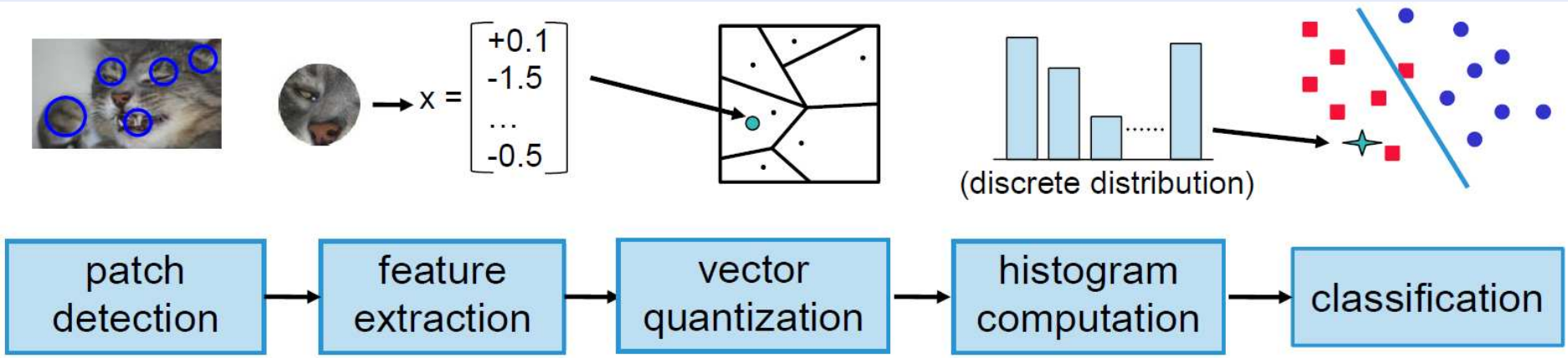
Action Classification – SVM RBF



Mean AUC: 63.66%.

The RBF σ was set to the minimum standard deviation of variables among the the training samples.

Ongoing work: Bags of Visual Words



- Using dense feature extraction around player positions for each action event:
- 5 scales $\times 9y \times 9x \times 9$ frames = 3645 feature vectors per player
- 1000 visual words

Ongoing/Future work

1. Player detection: use [Ramanan *et al.*, 2007] as a detector to replace heuristics.
2. Action recognition: evaluate 3D SIFT with BoF
3. Modules of the VAMPIRE system need improvement
4. Later: generalise the court detection and projection method
5. Maybe (or in parallel): investigate uses of gaze data as part of the loop

References


Kläser, A., Marszałek, M., & Schmid, C. (2008).
In: *British Machine Vision Conference* pp. 995–1004,.


Ramanan, D., Forsyth, D. A., & Zisserman, A. (2007).
IEEE Transactions on Pattern Analysis and Machine Intelligence, **29** (1), 65–81.


Walther, D. & Koch, C. (2006).
Neural Networks, **19**, 1395–1407.


Wang, H., Ullah, M. M., Käser, A., Laptev, I., & Schmid, C. (2009).
In: *Proc 20th British Machine Vision Conf, London, Sept 7-10* ,.