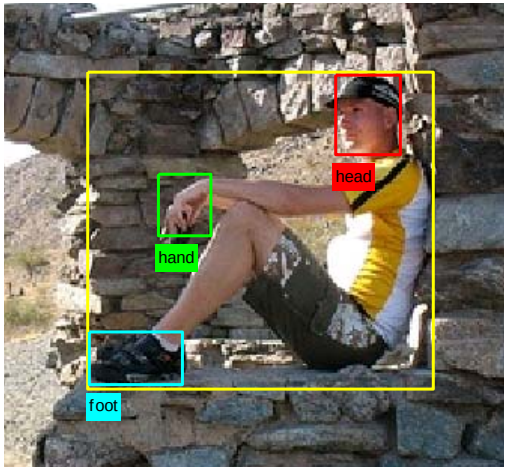
The PASCAL Visual Object Classes Challenge 2010 (VOC2010)

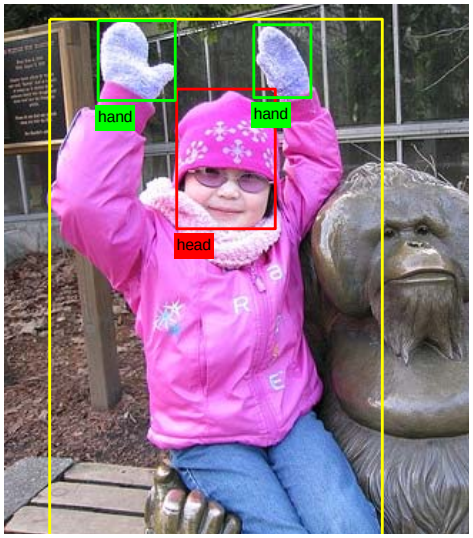
Part 5 – Person Layout Taster Challenge

Mark Everingham
Luc Van Gool
Chris Williams
John Winn
Andrew Zisserman

Person Layout Taster Challenge

- Given the bounding box of a person, predict the positions of head, hands and feet.





- Encourage research on more detailed image interpretation

Dataset Statistics

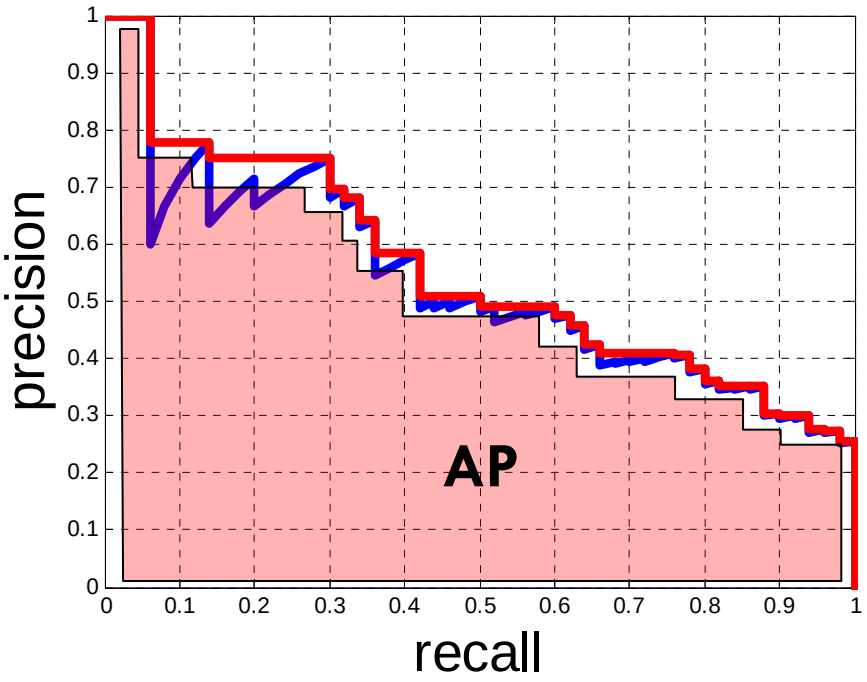
- Around 20% increase in size over VOC2009

	Training		Testing	
Images	376	(317)	320	(269)
Objects	576	(475)	505	(424)

VOC2009 counts shown in brackets

- Set of images taken (and removed) from main dataset
- Images contain only people (none of other 19 classes)

Evaluation



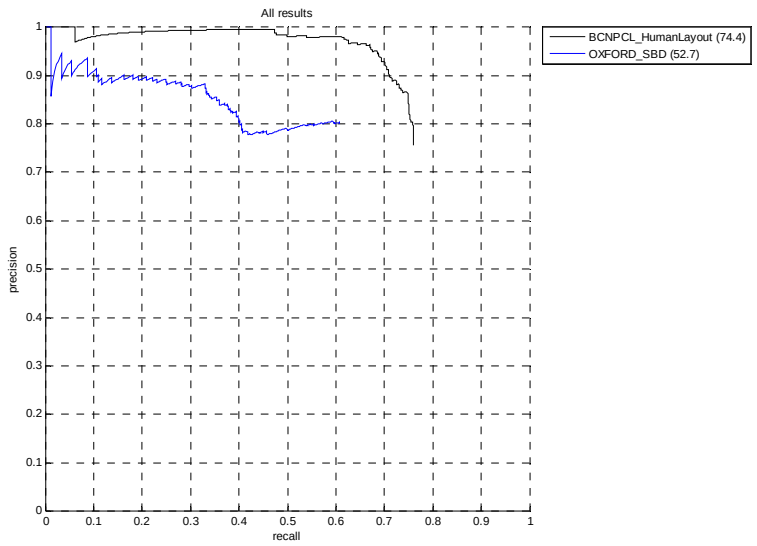
- 2010: Treated as three **separate** detection tasks: Head, Hands, Feet
- Evaluation by AP as in main detection task

- 2007-9 required correct prediction of set of parts visible and bounding boxes: not sensitive enough
- **Invitation: propose a better evaluation scheme!**

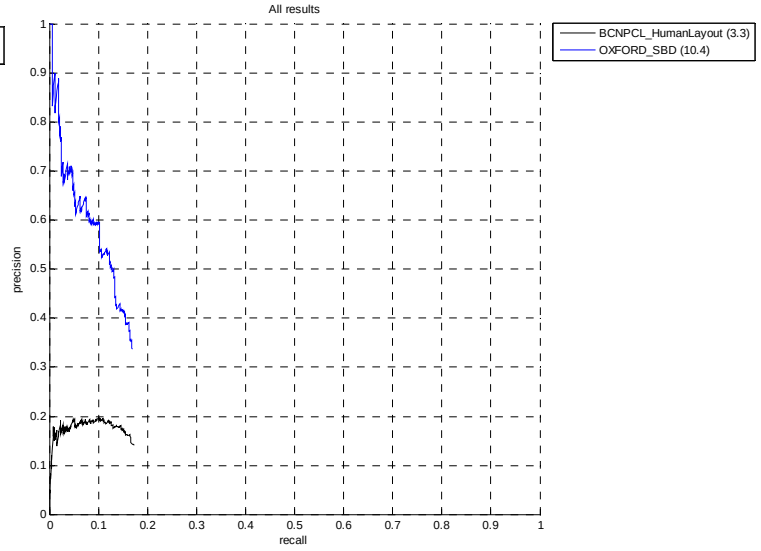
Results

- Two methods, both using detectors trained on other data
  - Oxford method does not attempt to detect feet

	Head	Hand	Foot
BCNPCL_HumanLayout	74.4	3.3	1.2
OXFORD_SBD	52.7	10.4	0.0



Head

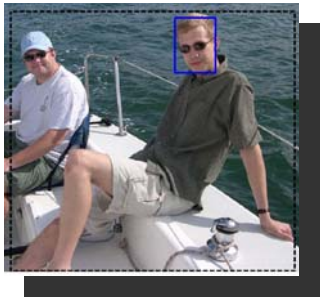


Hands

Combining detectors for human layout analysis.

M. Drozdzal, A. Hernández, S. Seguí, X. Baró, S. Escalera, A. Lapedriza, D. Masip, P. Radeva, J. Vitrià
Barcelona Perceptual Computing Lab
Universitat de Barcelona, Computer Vision Center, Universitat Oberta de Catalunya

Assumption: The head is visible and is the most reliable human part to detect. Head position is used as an anchor point for detecting other parts (hand and feet).



HEAD Average Precision %: 74.4

The head is detected by integrating evidences from several state-of-the-art part detectors: **frontal and profile faces** (OpenCV), **person parts** (*Felzenszwalb et al*), **informative poselets** (selected from the full set of *Bourdev at al*), and pictorial models (*Ramanan et al*.).

Integration is based on **weighted hierarchical clustering** of head windows hypothesis.

HANDS Average Precision %: 3.3

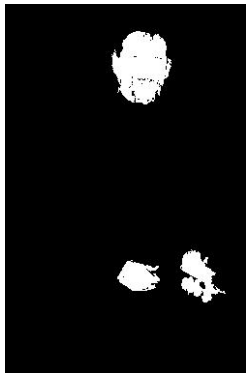
A specific **skin model** is extracted from the detected face. The image is segmented using Mean Shift with color information. Then, hand blobs are seeked using the learned skin model and specified geometric constraints.

FEET Average Precision %: 1.2

We use a person detection system proposed by *Felzenszwalb et al.* We have selected two 2 full body models which include feet in their components. For each model we choose the component related to the leg position. Inside the leg box the foot box position is learned by cross-validation.

OXFORD_SBD: Skin Based Layout Detection - I

1. Skin detection – learn image specific skin colour classifier
   - detect upper body and face (using latent SVM Felzenszwalb *et al.* detector)
   - detect skin pixels on face using a general skin classifier
   - learn image specific skin classifier from these pixels and their neighbours
   - classify image pixels to detect skin regions



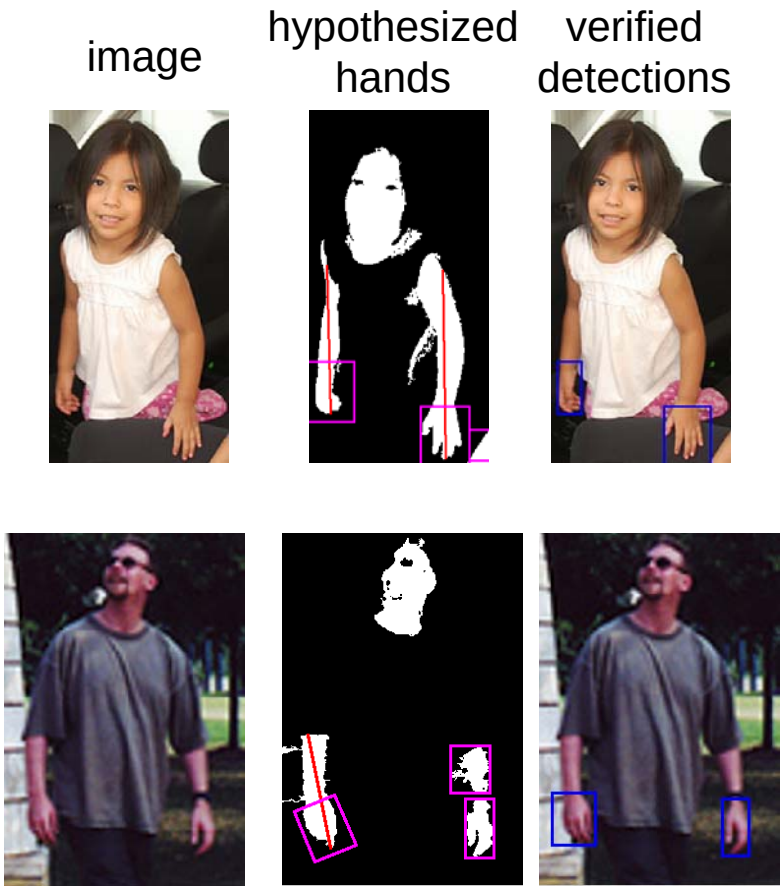
OXFORD_SBD: Skin Based Layout Detection - II

2. Hypothesize hands
   - on isolated blobs
   - at the end of arm shaped regions

3. Verify and score hands – using:
   - a hand classifier (RBF SVM on HOG)
   - upper body pictorial structure

4. Rank images based on hand score

Note, feet are not considered



Prizes

- Joint Winners:
  - **BCNPCL_Human_Layout**
    M. Drozdzal, A. Hernández, S. Seguí, X. Baró, S. Escalera, A. Lapedriza, D. Masip, P. Radeva, J. Vitrià
    Computer Vision Center Barcelona;
    Universitat Oberta de Catalunya
  - **OXFORD_SBD**
    Arpit Mittal, Andrew Zisserman, Philip H. S. Torr, Manuel J. Marin
    University of Oxford
    Oxford Brookes University