Combining detectors for human layout analysis.

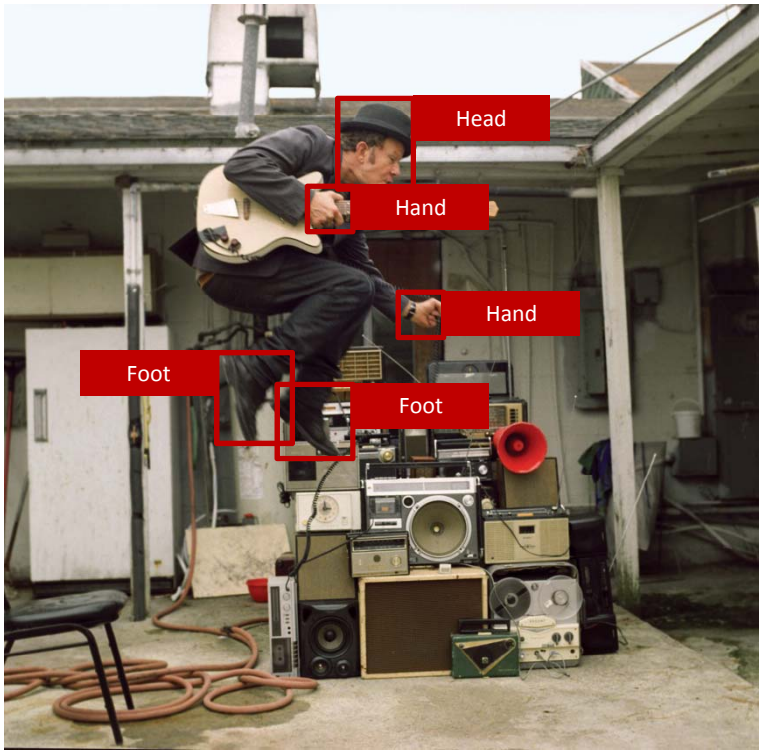
M. Drozdzal, A. Hernández, S. Seguí, X. Baró, S. Escalera, A. Lapedriza, D. Masip, P. Radeva, J. Vitrià

Barcelona Perceptual Computing Lab

Dept. de Matemàtica Aplicada i Anàlisi, Facultat de Matemàtiques
Universitat de Barcelona, Gran Via 585, 08007 Barcelona, Catalonia, Spain

Human Layout Analysis

Predicting the bounding box, label and presence of each part of a person (head, hands, feet)...



… in a dataset depicting extremely variable human poses and viewpoints in complex backgrounds.

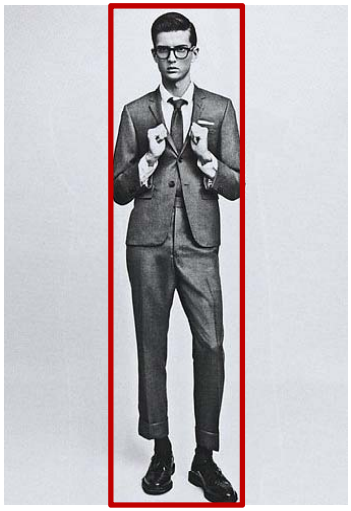
Approach

Our main objective was **not to build a new approach** to deal with the full complexity of the problem but to test some **state-of-art detector algorithms** in order to evaluate their limits.

We made the following assumptions:

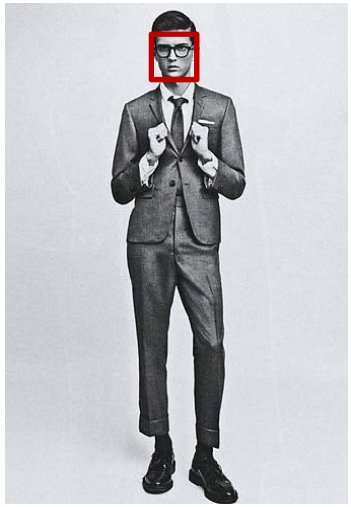
- We are provided with a bounding box of the visible parts of the body.

- The head is visible is the most reliable human part to detect. The head can be used as an anchor for other parts.

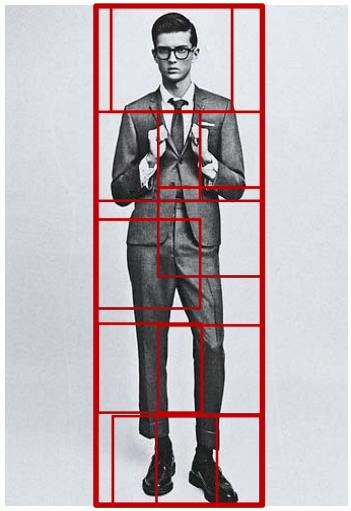


Head Detection

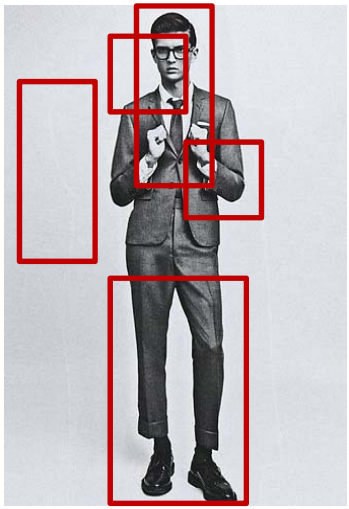
The head is detected by integrating several state-of-the-art part detectors:



Face (frontal + lateral) detection



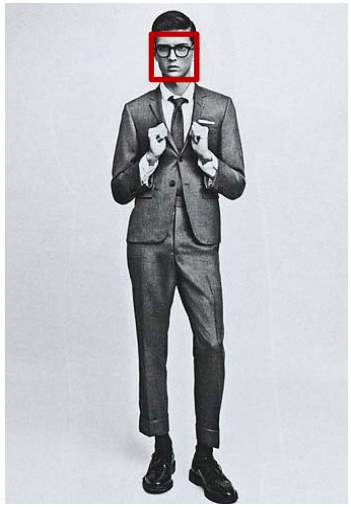
Person Detection using *Discriminatively Trained Part-Based Models*



Person detection using *poselets*

Person detection using Pictorial Model

Head Detection: Face

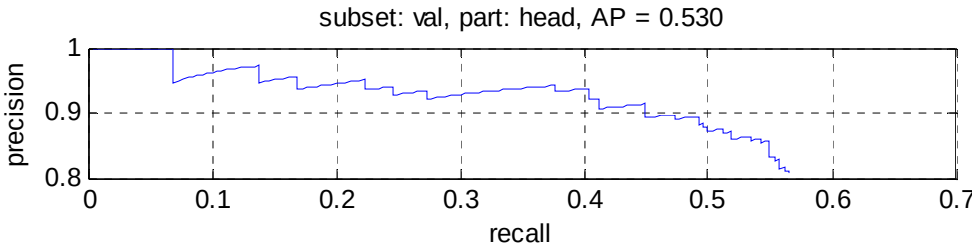


Face (frontal + lateral) detection

Faces were detected with OpenCV 2.1.

Details of the implementation:

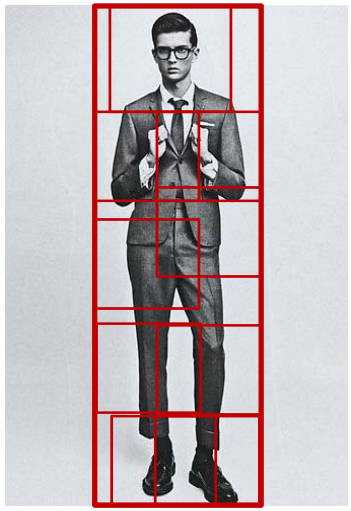
- We use the following cascades:
  - Frontal face (*default*, *alt*, *alt2*, *alt_tree*).
  - Lateral face (*profile*).
- Each cascade return several (from 0 up to N) hypothesis about head position.
- To integrate the results we use hierarchical clustering.
- The final head box is the one with the maximum score given by hierarchical clustering.



References: Viola, Jones: **Robust Real-time Object Detection**, IJCV 2001

Head Detection: Body (i)

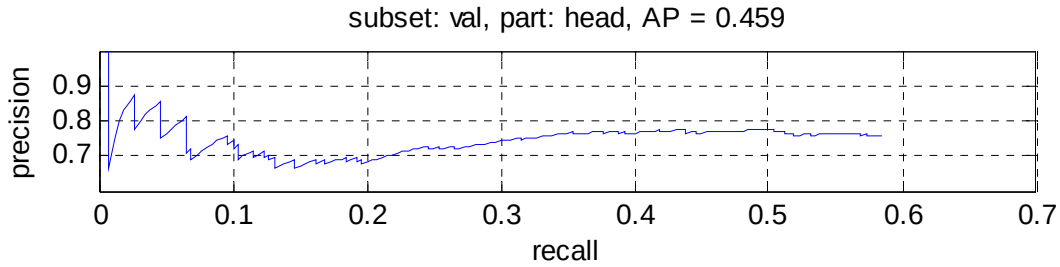
We use a person detection system proposed by *Felzenszwalb et al.*



Person detection

Details of the implementation:

- Software version: **Discriminatively Trained Deformable Part Models Version 4.**
- Based on model aspect analysis we choose 4 models which best detect the head position.
- For each model we choose the component related with head position in order to fix the box.

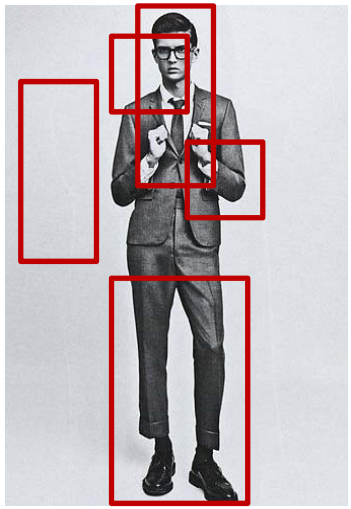


References:

- *P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan,* **Object Detection with Discriminatively Trained Part Based Models,** PAMI 2009

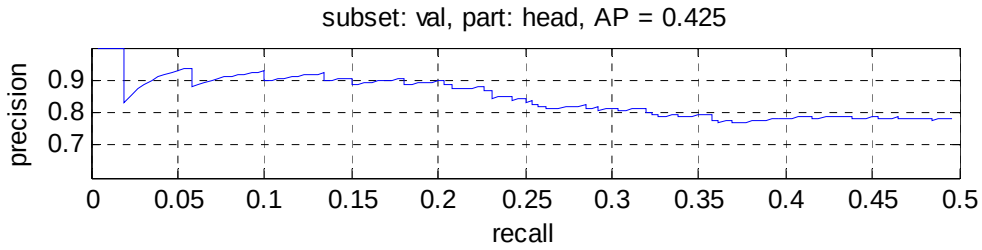
Head Detection: Body (ii)

We use the body detection system proposed by *Bourdev et al.*



Poselets detection

- Initially, we used the set of 1138 *poselets* trained from the H3D database.
- The poselets were trained to vote for position and size of the head.
- In order to improve results a hierarchical clustering **per poselet** was introduced.
- From original poselets set, **we selected** the 239 poselets which gives the best, in terms of reliability, votes for the head position. The used selection criteria was the standard deviation (*std*) of votes for head.
- If *std* was smaller than a defined threshold then the poselet was defined as reliable.



Reference:

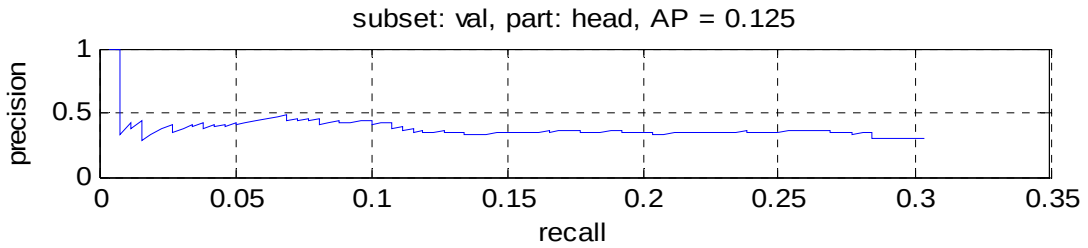
- Lubomir Bourdev, Jitendra Malik, **Poselets: Body Part Detectors Trained Using 3D Human Pose Annotations**, ICCV 2009.

Head Detection: Body (iii)

Pictorial Model

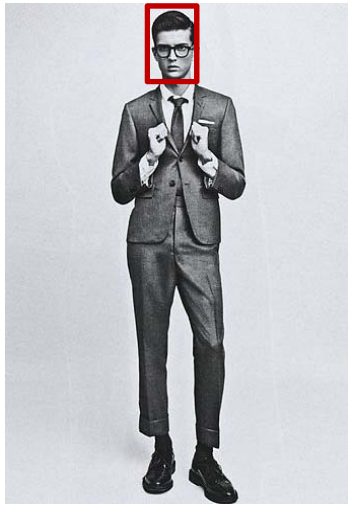
We used the implementation of CRF-based pose recovery.

The method is applied at different scales and the solution with the maximum response is selected.

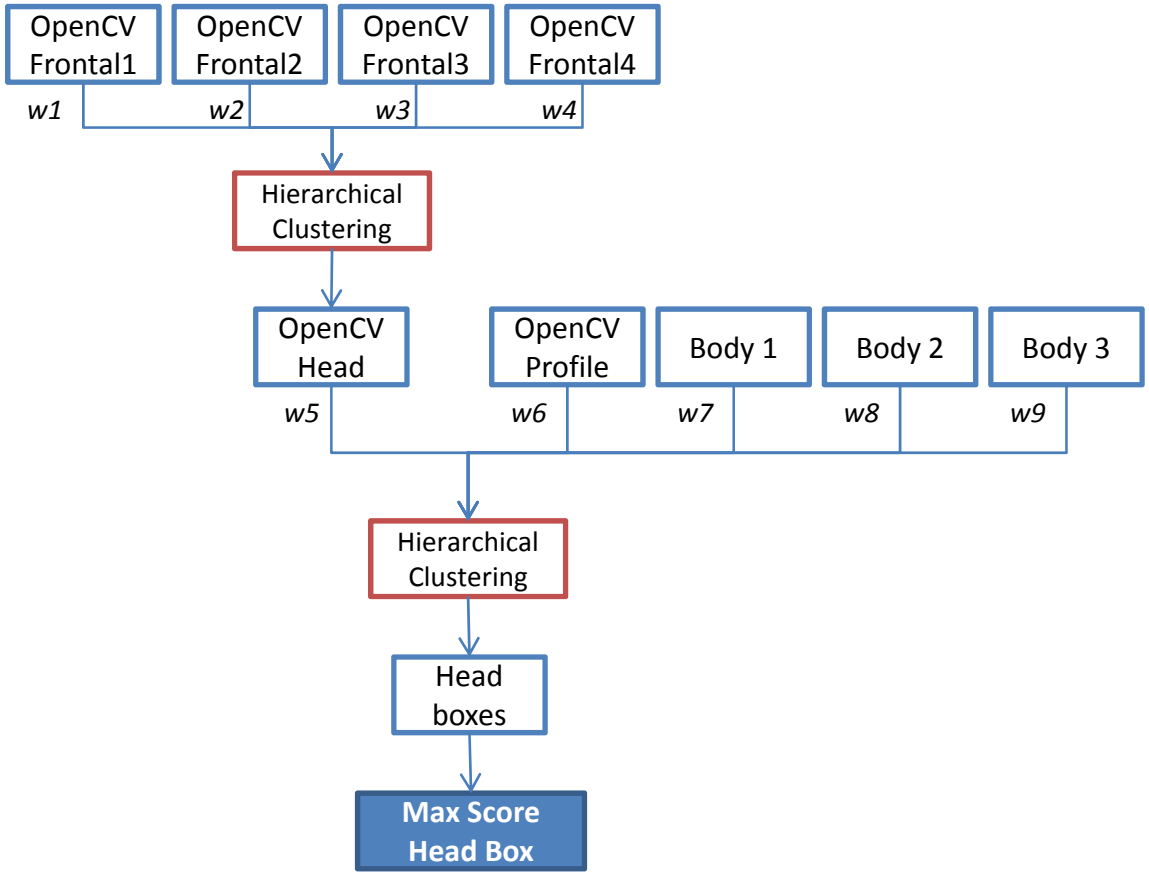


Reference:
• Ramanan, D. **Learning to Parse Images of Articulated Bodies** , *In NIPS, 2006*.

Head Detection: Integration

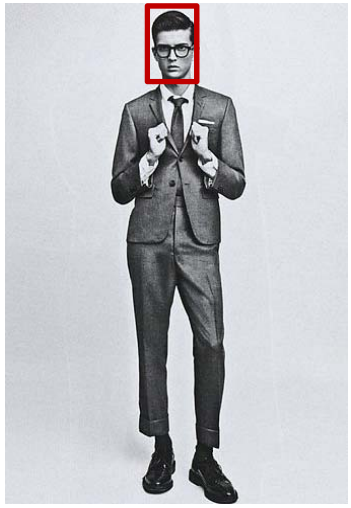


HEAD



wi are the weights assigned to each detector. These weights are set using cross-validation.

Head Detection: Integration



HEAD

For each head box the hierarchical clustering computes the score.

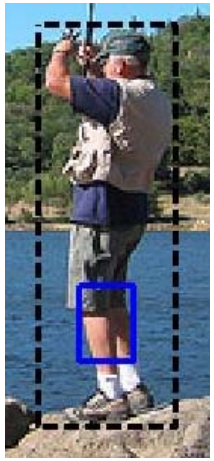
The higher score the larger probability of correct head detection.

The max score head box is the head detection with the biggest probability.

The score is used as a confidence value for each detection.

If there are 2 highly probable heads inside layout bounding box the score is divided by 2.

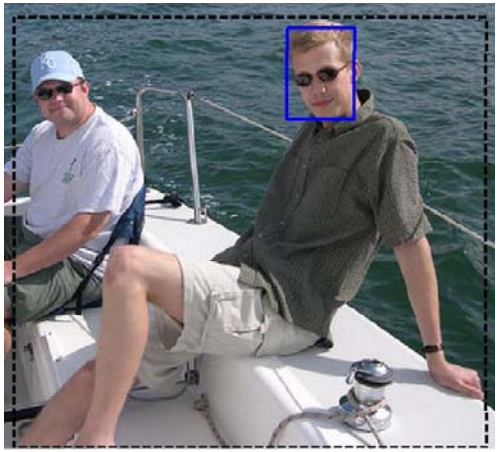
Head Detection: Results



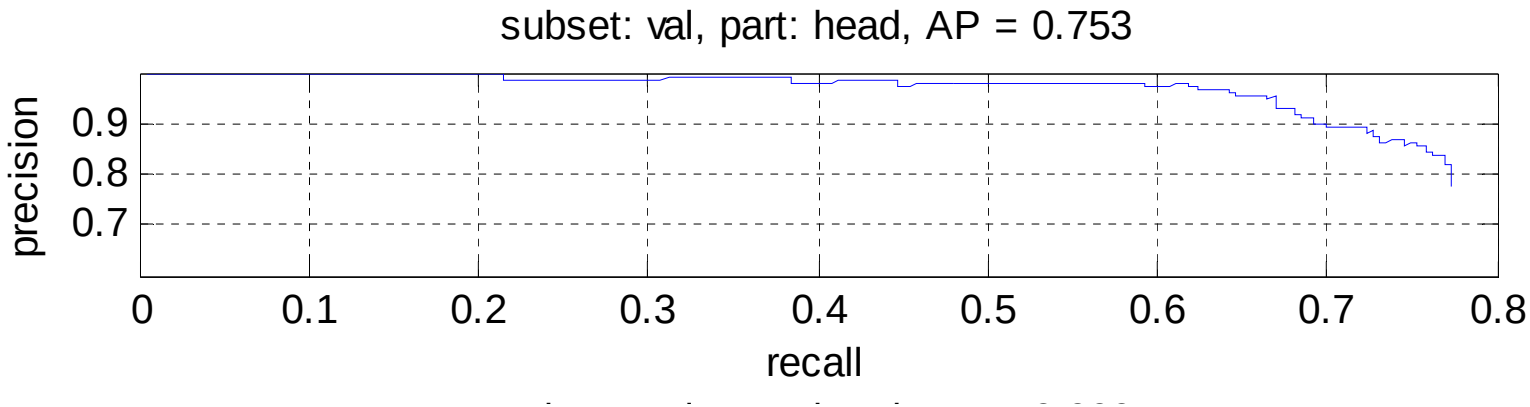
Confidence 0.5

Confidence 0.8

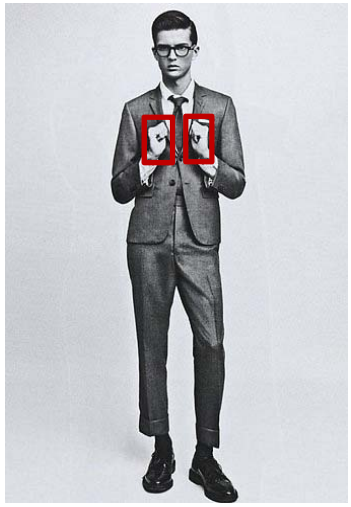
Confidence 1.6



Confidence 2.25



Hand Detection



Hand Detection

Image

Head Detection

Face Skin Segmentation

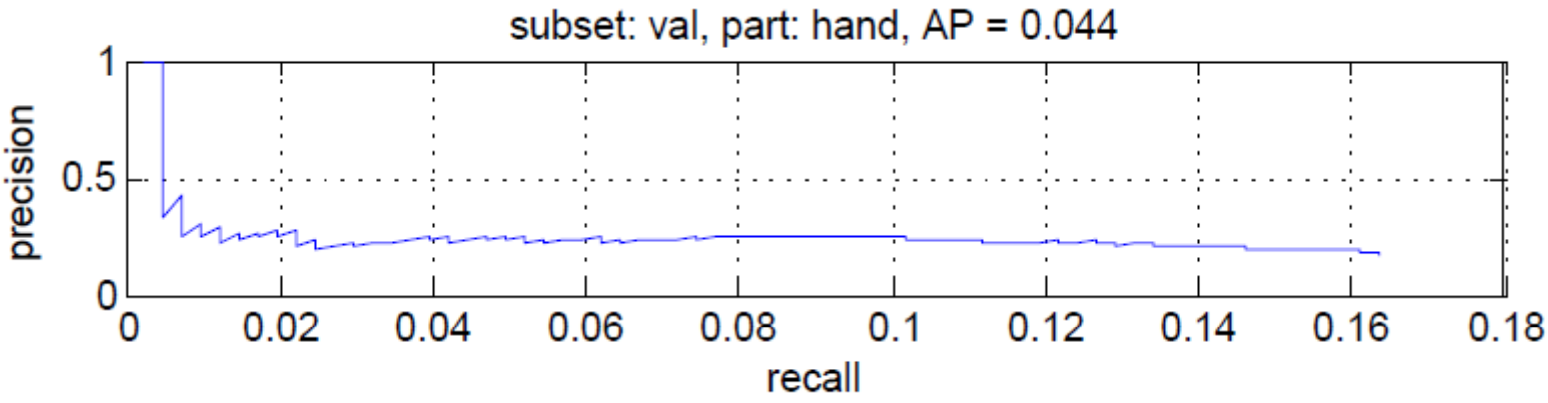
Image Segmentation using [1]

Blob selection using the person skin model

Filtering

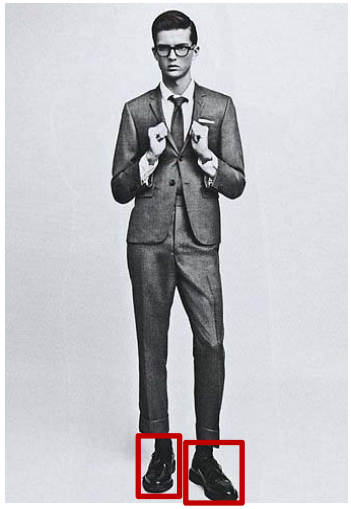
References:
[1] Dorin Comaniciu , Peter Meer , Senior Member, **Mean shift: A robust approach toward feature space analysis ,** PAMI 2002

Hand Detection: Results



Feet Detection

We use a person detection system proposed by *Felzenszwalb et al.*



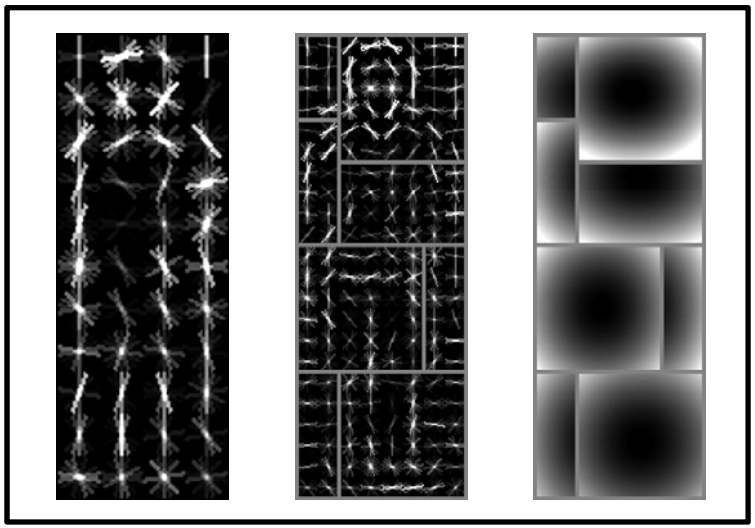
Feet Detection

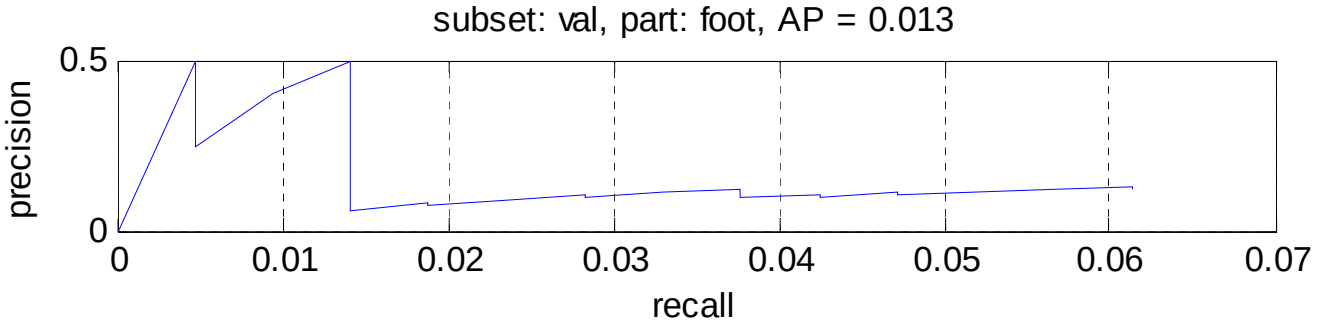
Details of the implementation:

- Software: **Discriminatively Trained Deformable Part Models Version 4.**
- Based on model aspect analysis we choose 2 full body models which detect the feet.
- For each model we choose the component related with lower leg position.
- Inside lower leg box the foot box is fixed by cross-validation.

References: *P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan,* **Object Detection with Discriminatively Trained Part Based Models,** PAMI 2009

Feet Detection: Results



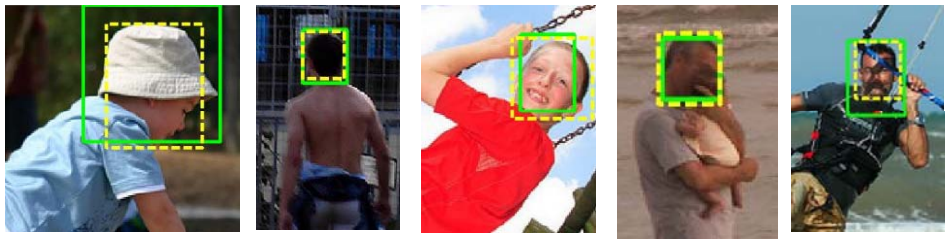


Conclusions

By integrating state-of-the-art part detectors is possible to obtain good results in **head detection**.

Our detector system allows to detect heads in following conditions:

- Frontal Heads
- Picture taken from behind
- Heads with partial occlusions



Our detector has problems with :

- Rotated images.
- Profile heads with no standard pose.
- Several heads in the layout box.

Head detection is mature enough for developing a fast, robust detector, provided that we find how to efficiently deal with these problems.

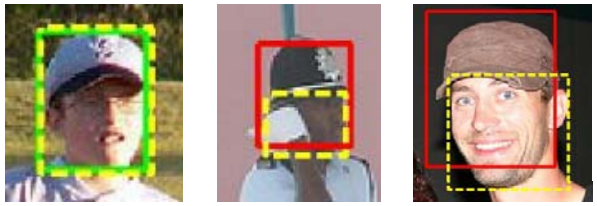
Conclusions

Feet detection could be improved by training more **diverse models for legs**. Presence is an open problem for low resolution images.

Hand detection should be approached as **a search problem** (f.e. using the CHAINS model [2]). These models could be easily extended to feet detection.

Detectors must be integrated in **pictorial models** that are able to deal with a large variety of detectors.

It is necessary to build better **databases** with clear labeling policies.



References:
[2] L Karlinsky, M Dinerstein, D Harari, S Ullman, **The chains model for detecting parts by their context ,** CVPR 2010

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.

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. This model is used to segment the image using Mean Shift. Resulting blobs are filtered using 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.