Data-Driven Parsing by Matching

Daphne Tsatsoulis and David Forsyth
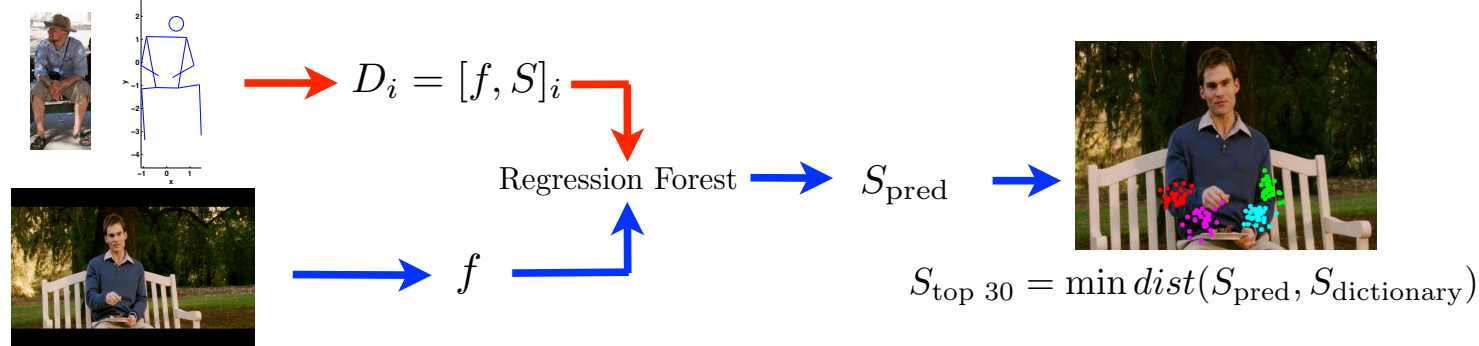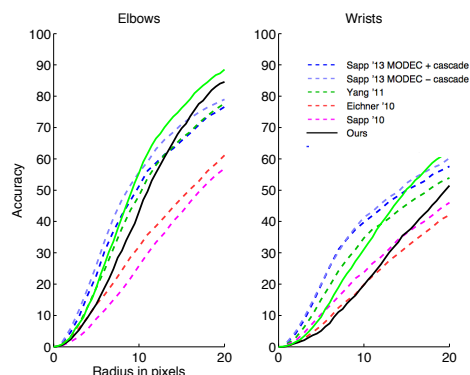{tsatsou2, daf}@illinois.edu

Motivation

Human body parsing involves taking an image that is known to contain a person and reporting locations of body parts. It remains a difficult problem to produce fast, accurate parses. Our method approaches parsing differently from current methods and is simple and effective. A significant novelty of our method is that it naturally produces 3D estimates of wrist and elbow position relative to the torso and in dependent of view direction.

Method

Our method is straightforward and has accuracy on standard datasets comparable to state of the art. Unlike state of the art our method does not search for individual body parts or poselets. Instead, it regresses image features to joint locations in a normalized torso-centric space. Our method then matches predicted poses to a dictionary of real poses. All predictions made are of natural upper-body parses that can be 2- or 3-dimensional.

$$D_i = [f, S]_i$$

Regression Forest $\rightarrow S_{\text{pred}} \rightarrow$

f

$$S_{\text{top 30}} = \min dist(S_{\text{pred}}, S_{\text{dictionary}})$$
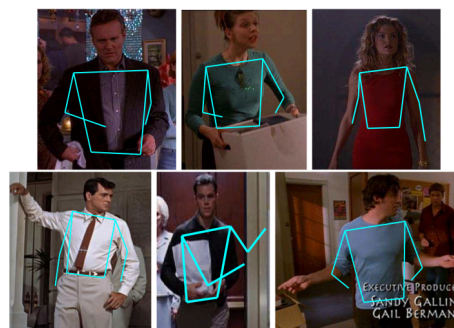
Quantitative Results

Radial precision plots of elbow and wrist joint predictions in 2D. Radius (r) is reported for a normalized torso diagonal of 100 pixels. The black line is our top 1 match. The green line is the best result of our top 30 matches.
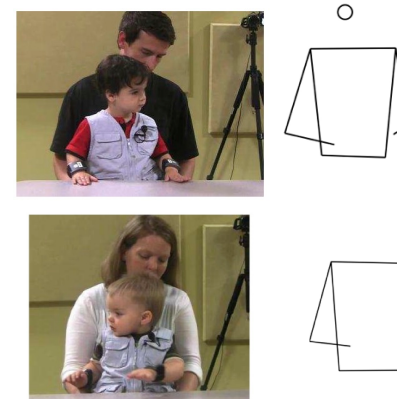
3D Results

We can predict 3D skeletons by building a dictionary of 3D skeletons.

Real Results

Examples of common errors in our system. When our system fails it produces output that is still a natural skeletal configuration unlike other parsers.

New Datasets

We can predict results for datasets we did not train on.