Supervised Training and Computer Vision

Ali Eslami

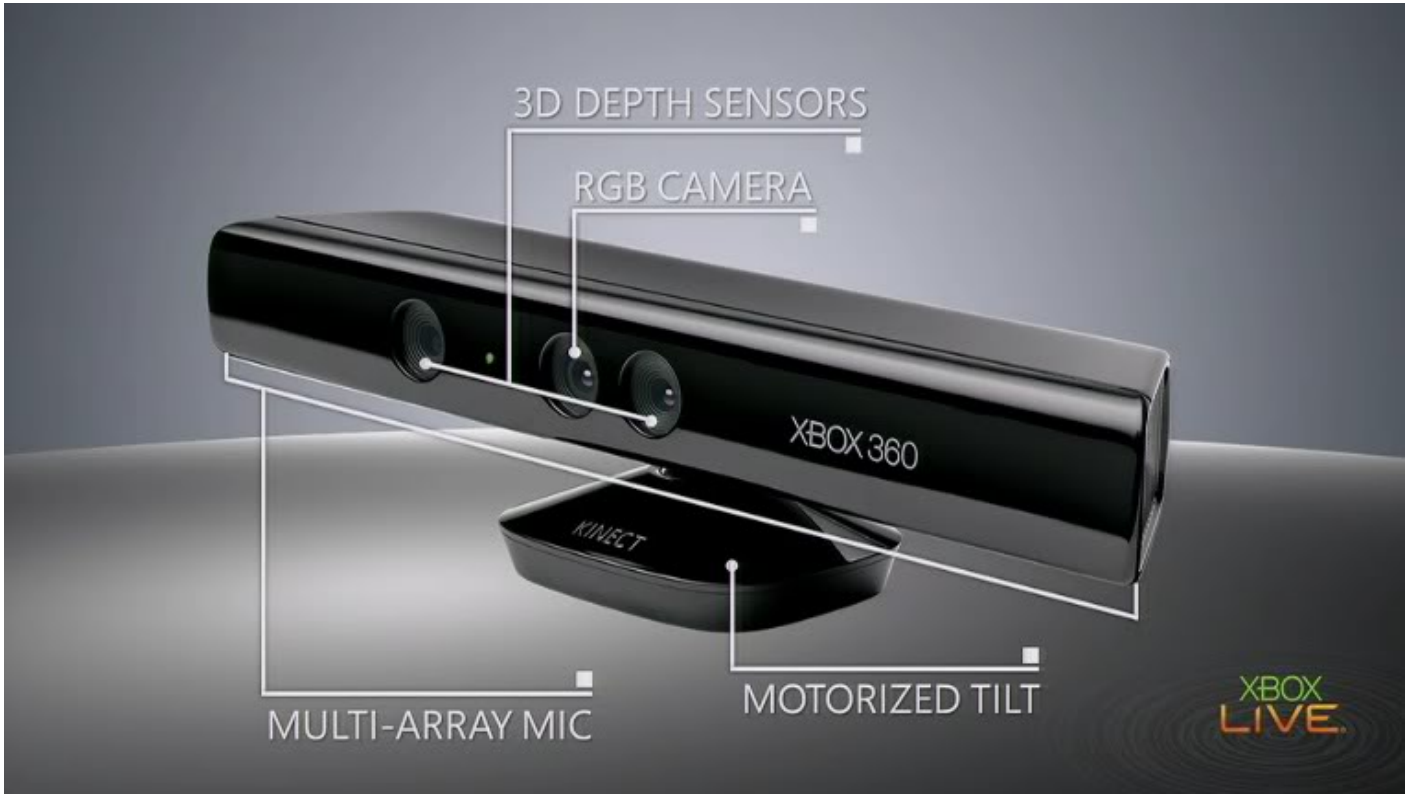
THE UNIVERSITY of EDINBURGH
informatics
ianc | Institute for Adaptive and Neural Computation

Brainstorm Coffee
May 18, 2011

Case Study

The Setting: Kinect

The Setting: Kinect



The Setting: Kinect

http://youtube.com/watch?v=So6AjnKPr5A

The Setting: Kinect

Requirements:

- **Performance**: to run at around 200 frames/second.
- **Accuracy**: to compete with rival hardware.
- **Robustness**: for people of varying shapes and sizes.

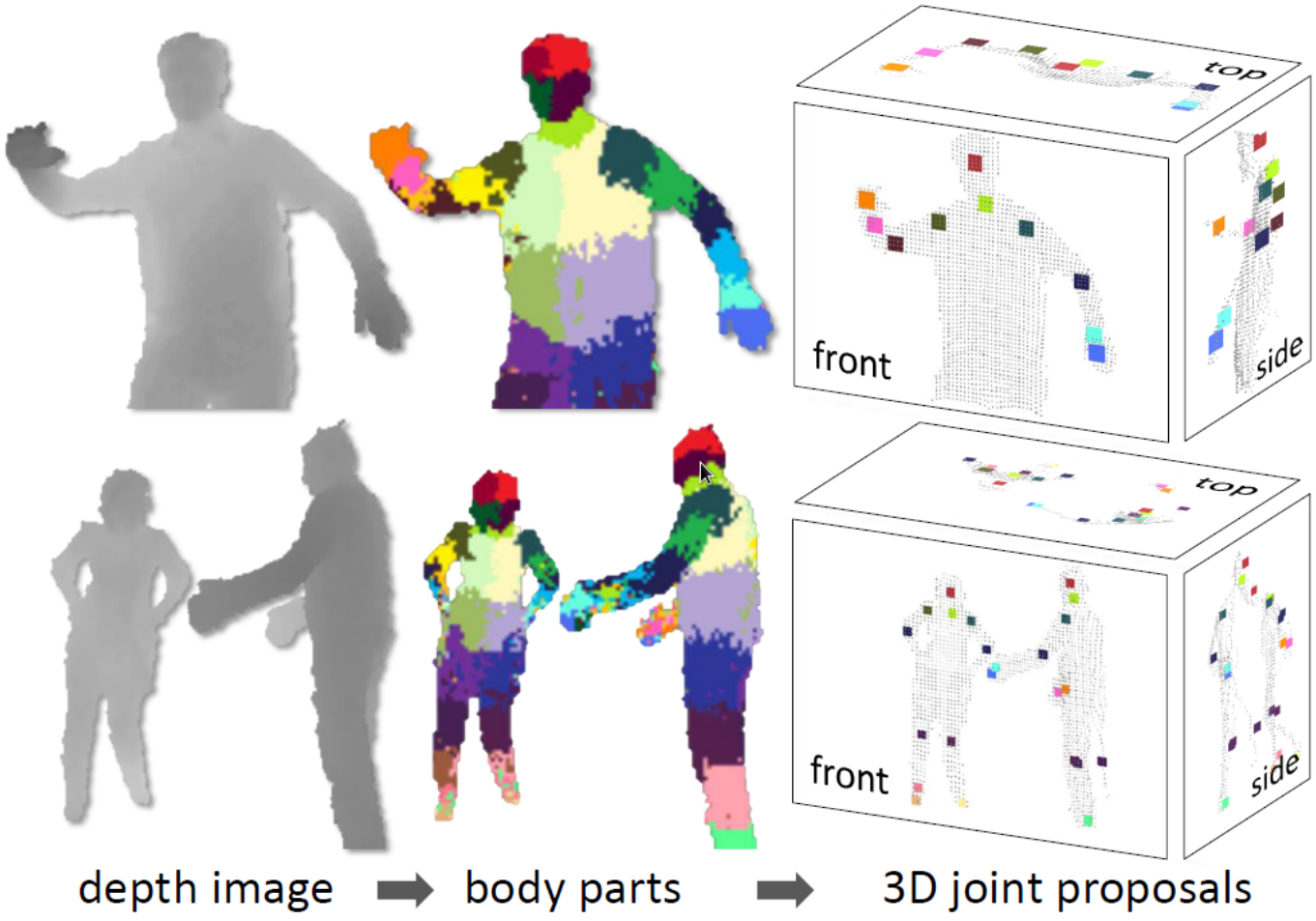
A Solution

Real-Time Pose Recognition in Parts from Single Depth Images

Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman and Andrew Blake

CVPR 2011

Pipeline


top
front
side
top
front
side
depth image
body parts
3D joint proposals

Approach

Model

Simple:

- Randomized decision forests.
- Pixel-per-pixel.
- Frame-per-frame.

Training

- **Heavily** supervised.
- **Huge** training dataset.

Features

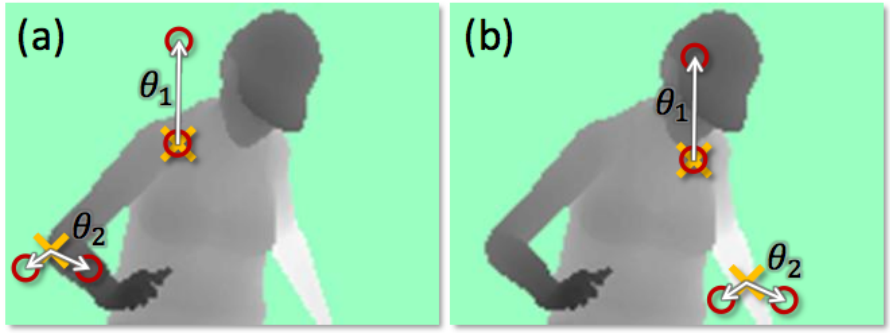


Figure 3. **Depth image features.** The yellow crosses indicates the pixel $\mathbf{x}$ being classified. The red circles indicate the offset pixels as defined in Eq. 1. In (a), the two example features give a large depth difference response. In (b), the same two features at new image locations give a much smaller response.

$$f_\theta(I, \mathbf{x}) = d_I\left(\mathbf{x} + \frac{\mathbf{u}}{d_I(\mathbf{x})}\right) - d_I\left(\mathbf{x} + \frac{\mathbf{v}}{d_I(\mathbf{x})}\right)$$

Forests

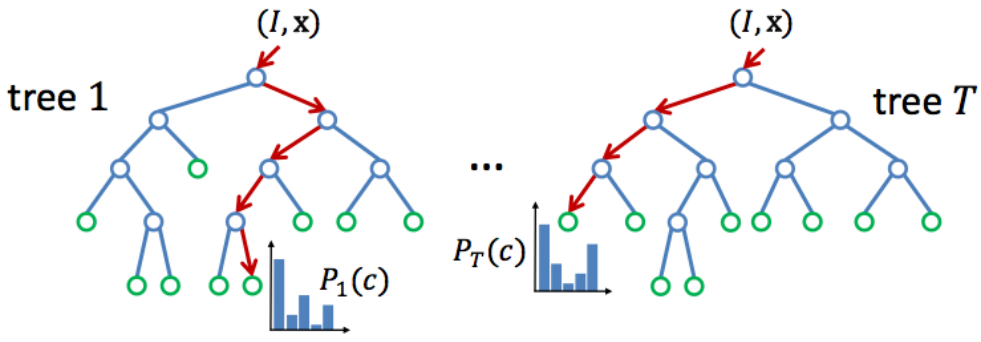


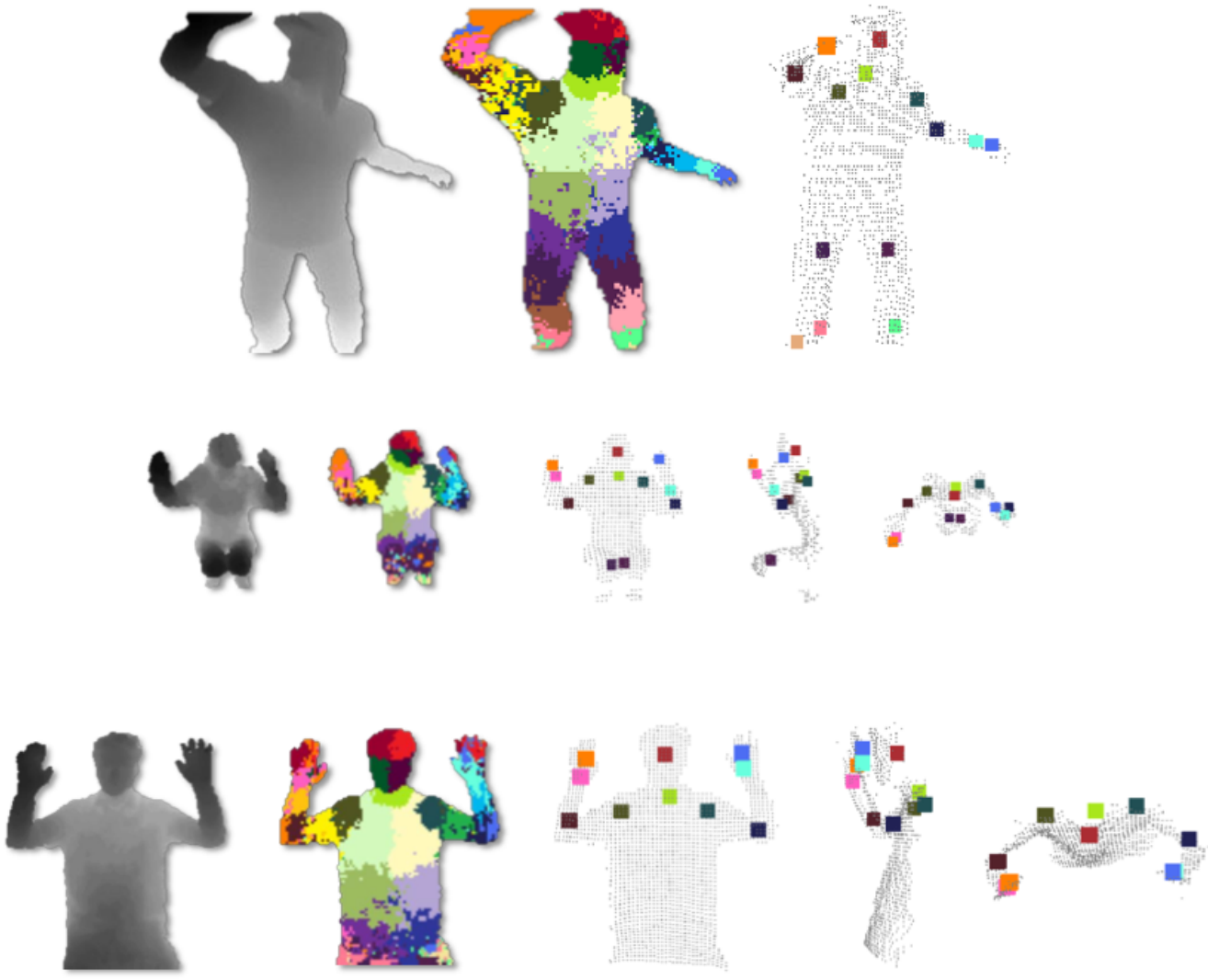
Figure 4. **Randomized Decision Forests.** A forest is an ensemble of trees. Each tree consists of split nodes (blue) and leaf nodes (green). The red arrows indicate the different paths that might be taken by different trees for a particular input.

Training data

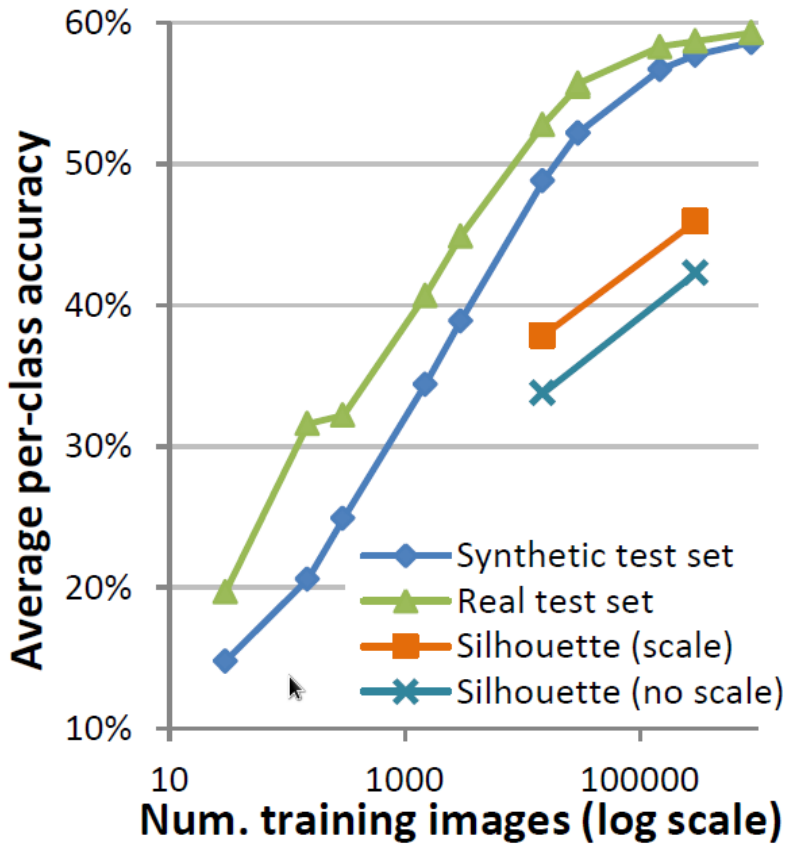
- Entirely synthetic.
  - Mocap data for pose.
  - 15 base meshes for shape.
  - Render w/ texture map for ground truth.
- Total of **900,000** training images.

Training data

Sample results



Number of training images



Food for thought

Question 1

"Why is it that we don't **automatically** generate **heavily annotated** training datasets for more machine learning problems?"

Question 1

"Why is it that we don't **automatically** generate **heavily annotated** training datasets for more machine learning problems?"

This should theoretically be possible for any[?] task that humans can perform well without the use of computers.

Question 2

"Is there much point to an unsupervised algorithm, that can potentially train on an unlimited amount of data, when the extra data doesn't make it any better at its job?"

Question 2

"Is there much point to an unsupervised algorithm, that can potentially train on an unlimited amount of data, when the extra data doesn't make it any better at its job?"

- Usually the job is not to 'learn from lots of unlabelled data'.
- Case for nonparametric models?

Question 3

"Are **learning** algorithms the same as the **thinking** systems of the previous century?"

Question 3

"Are **learning** algorithms the same as the **thinking** systems of the previous century?"

Perhaps once we see past the *coolness* of learning, we can spend our time making systems that **actually work**?

Discussion