Kinect Case Study

CSE P 576
Larry Zitnick (larryz@microsoft.com)

http://www.youtube.com/watch?v=dTKINGSH9Po&feature=related
Questions

• Why a dot pattern?
• Why a laser?
• Why only one IR camera?
• Is the dot pattern random?
• Why is heat a problem?
• How is it calibrated?
• Why isn’t depth computed everywhere?
• Would it work outside?

Pose recognition

• Research in pose recognition has been on going for 20+ years.

• Many assumptions: multiple cameras, manual initialization, controlled/simple backgrounds
Kinect

- Why does depth help?

Algorithm design

Shotton et al. proposed two main steps:

1. Find body parts
2. Compute joint positions.

Real-Time Human Pose Recognition in Parts from Single Depth Images
Jamie Shotton Andrew Fitzgibbon Mat Cook Toby Sharp Mark Finocchio
Richard Moore Alex Kipman Andrew Blake, CVPR 2011
Finding body parts

- What should we use for a feature?
- What should we use for a classifier?

Finding body parts

- What should we use for a feature?
  - Difference in depth
- What should we use for a classifier?
  - Random Decision Forests

Features

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

Classification

\[ P(c|I, x) = \frac{1}{T} \sum_{t=1}^{T} P_t(c) \]

Learning:
1. Randomly choose a set of thresholds and features for splits.
2. Pick the threshold and feature that provide the largest information gain.
3. Recurse until a certain accuracy is reached or depth is obtained.
Implementation details

- 3 trees (depth 20) *(why so few?)*
- 300k unique training images per tree.
- 2000 candidate features, and 50 thresholds
- One day on 1000 core cluster.
- *Why RDF and not AdaBoost, SVMs, etc.?*

Synthetic data

Synthetic training/testing

Real test
Joint estimation

- Apply mean-shift clustering to the labeled pixels. (why mean shift?)
- “Push back” each mode to lie at the center of the part.

Failures

- Why would the system fail?
Video


Story about the making of Kinect: