Week 3

Stephan Boyer
UCF REU

The Project

- We seek to answer the question:

  "How many frames does it take to identify a human gesture from a depth video?"

- There are 12 gestures we are trying to recognize (J. LaViola, J. Norton, C. Wingrave).
- We are constructing an entire dataset from scratch.

The Goal

- To maximize the response time of depth-based video game systems like Microsoft's *Project Natal*.
- Depth-based systems won't be able to compete with normal button-based games if they aren't responsive enough.
- Video games need to be able to respond to a player's actions in real time (ideally, < 100 ms).
- 100ms * 1 sec/1000 ms * 30 frames/sec

    = **3 frames** ← {bears a small similarity to the one-shot project}

Dataset Construction

- I've developed a toolset of three applications:
  - "sp" allows me to test my cameras and view two live video streams side by side.
  - "ccw" allows me to calibrate the cameras by finding a homography from the left image to the right in "special circumstances" (when everything is far away from the camera and the stereo separation is small).
 - "dr" allows me to record depth videos.

The Algorithms

- "ccw" provides depth maps in real time to assist with calibration. This uses the faster **block matching** algorithm.

- "dr," which is used to produce the datasets, uses the slower, but less noisy **graph-cut** algorithm because it does not need to operate in real time.

- Both of these algorithms are provided by OpenCV.

Block Matching Results

Graph Cut Results
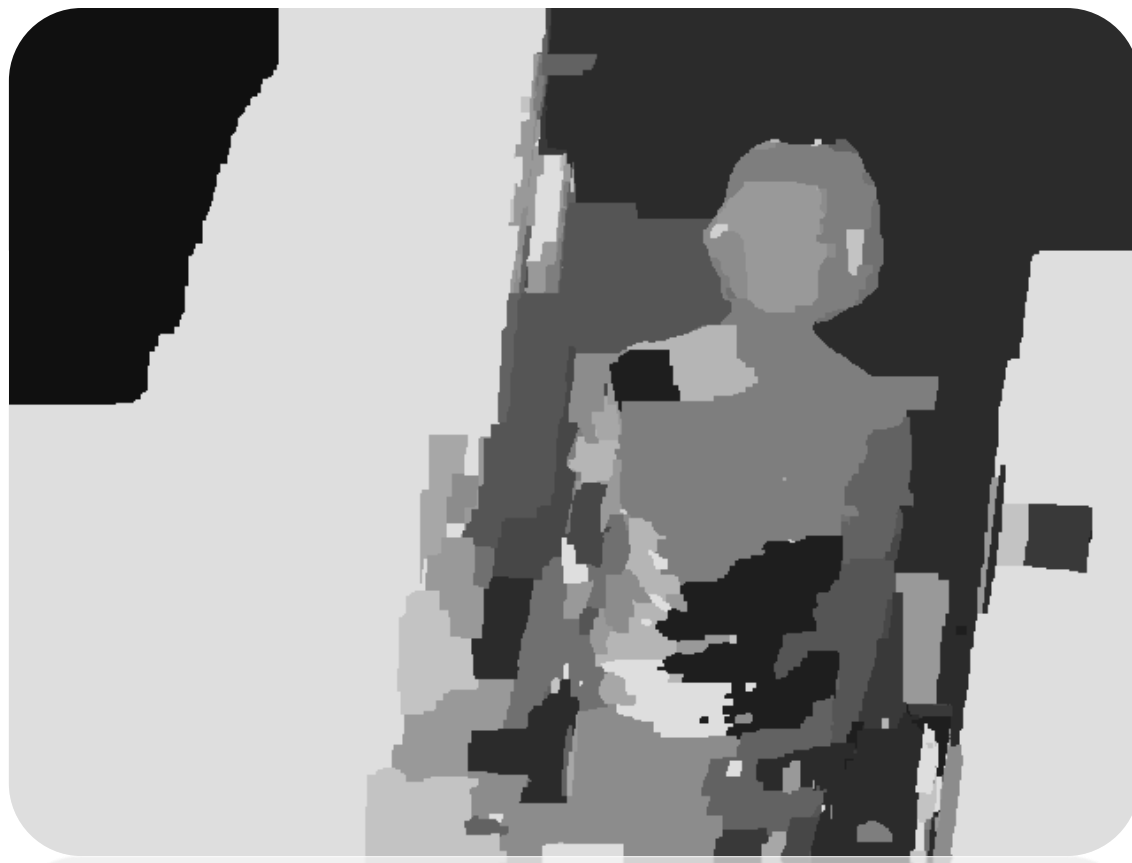
Future Plans

- Finish the dataset
- Learn more about the "bag of words" model (feature detection/representation, codebook generation, Bayseian models, Discriminative methods, etc.)
- Write an prototype algorithm for identifying the 12 human gestures
- Train the algorithm if necessary using the dataset
- Improve the algorithm by minimizing the number of frames required for recognition
  - Incorporation of one-shot techniques?
- Incorporate depth AND color?
- Compare results to normal color-only systems