Weeks 9, 10, 11

Stephan Boyer
UCF REU

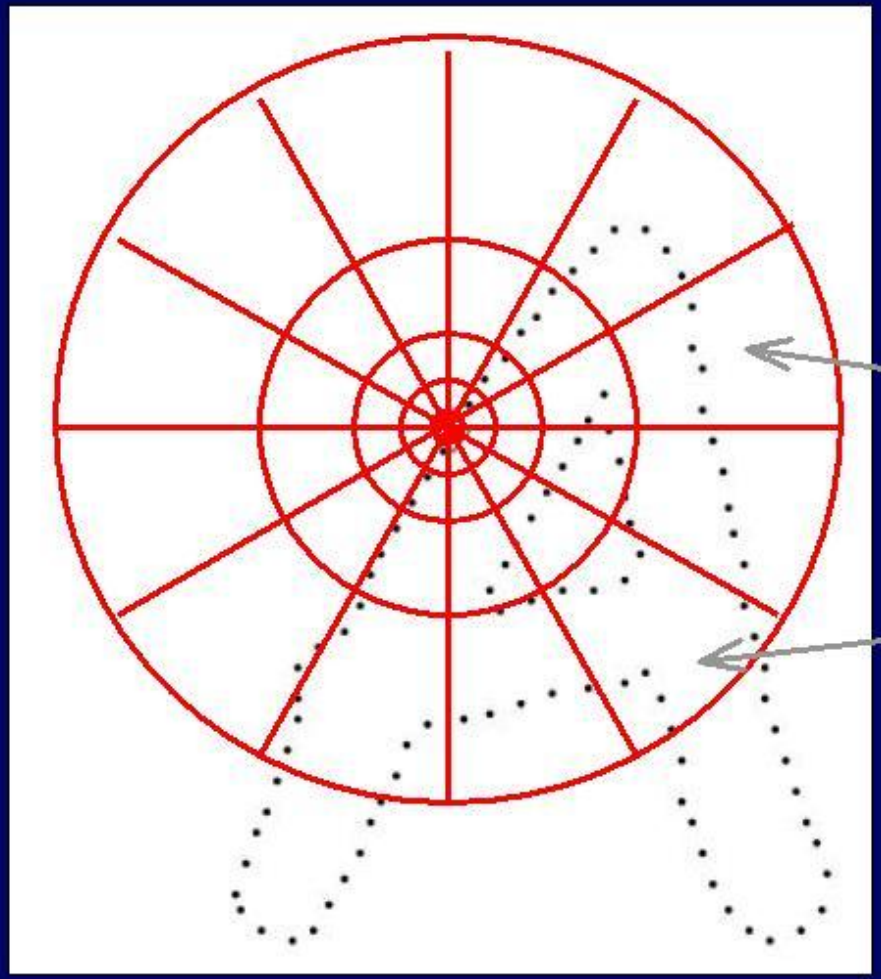
Since Last Time…

- Finished collecting the dataset! ☺
  - 10 different actors
  - 400 sequences
  - 200 transitions
  - Each gesture was recorded 40 times
  - Total of 2,400 videos
- Improved the postprocessing
- New grad student: Brian

Since Last Time…

- Implemented a simple feature descriptor based on shape context, modified to incorporate temporal information
- Implemented a second feature descriptor based on shape context alone (no temporal information)
- Results so far:
  - Weizmann dataset: 84.4% Accuracy
  - Boyer dataset: 48.2% Accuracy

Shape Contexts



Count the number of points inside each bin, e.g.:

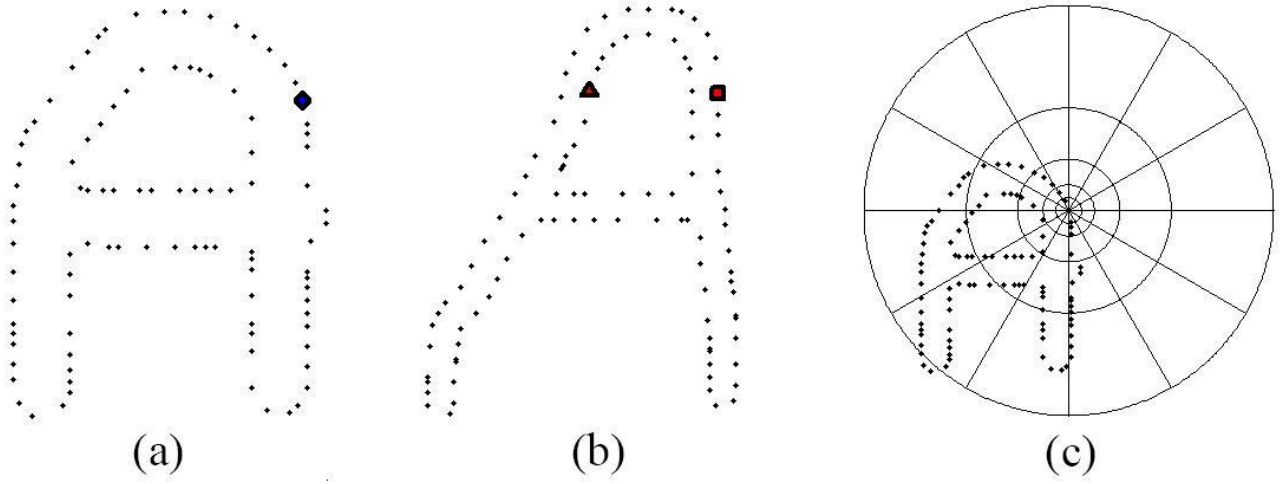
Count = 4

⋮

Count = 10

☞ Compact representation of distribution of points relative to each point

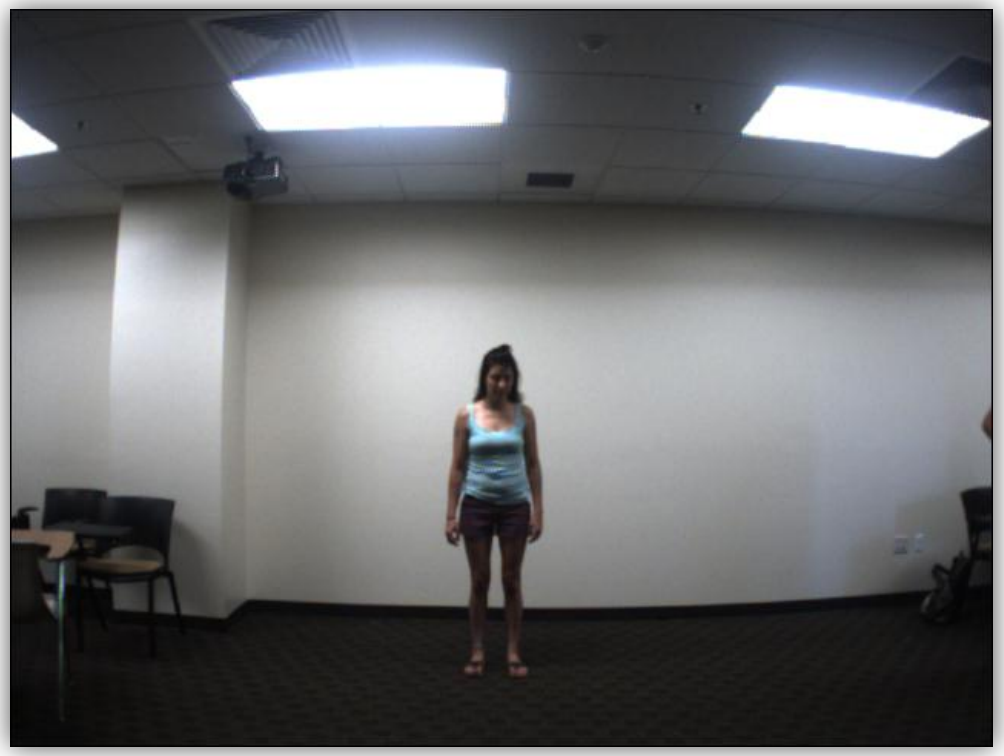
Shape Contexts



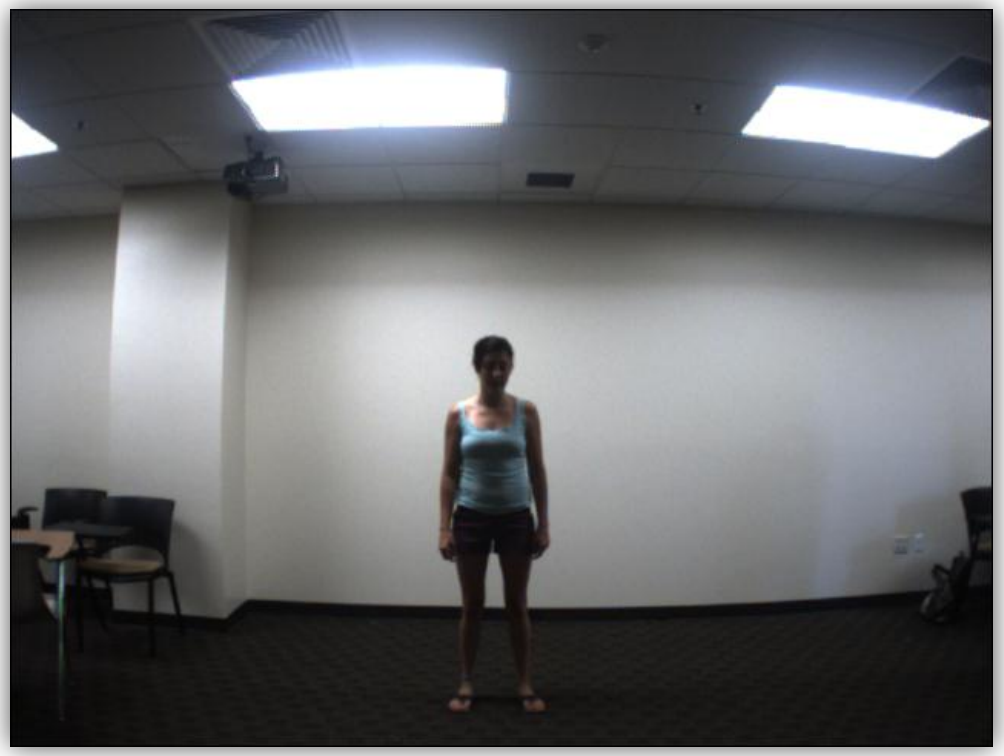
Problems Encountered

- There was little time to train the actors
- The recording environment was not optimal; side effects included leg amputations among other things. Reasons:
  - The color of the floor was apparently the same as the color of human legs
  - I noticed foreign elbows in some shots
  - Someone moved their backpack into the scene during a recording session (this interferes with background subtraction)
  - Chairs were also moved around
  - Stray shadows were cast into the scene
  - Doors opened and closed during recording (changing the lighting of the scene)

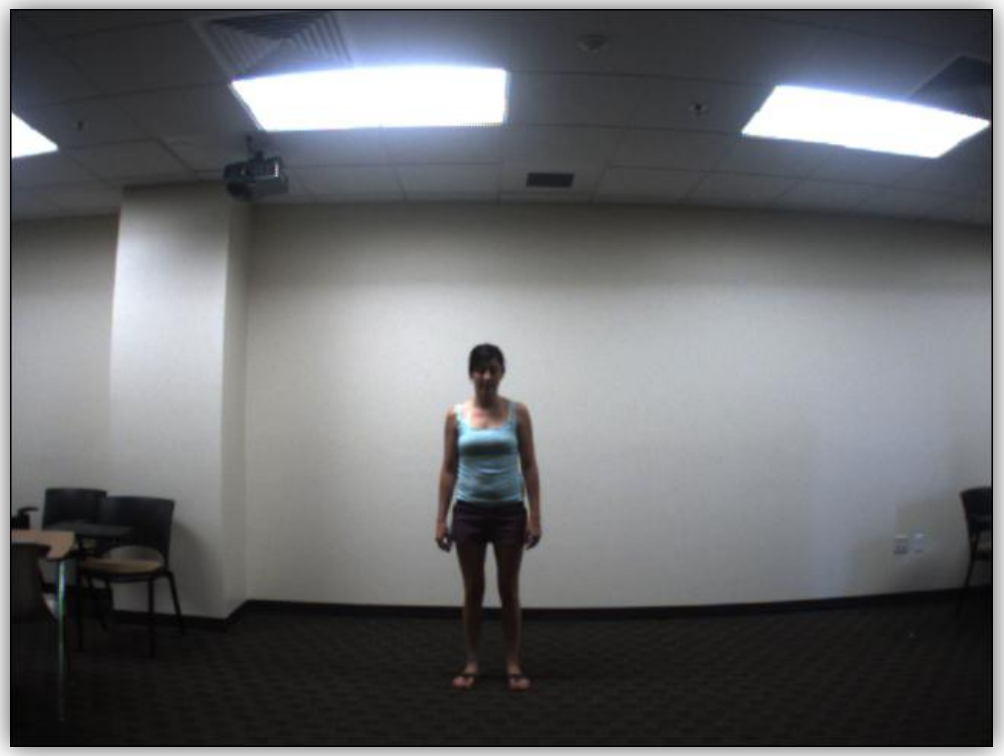
Foreign Elbows



Moving Backpacks and Chairs



Moving Backpacks and Chairs



Amputations

Confusion Matrix (Boyer)

30	6	4	0	2	3	0	5	2	0
3	21	7	9	1	2	0	8	0	0
1	1	16	4	3	2	1	0	1	0
0	2	4	13	0	0	0	2	5	0
2	3	3	0	15	5	0	2	7	7
0	0	0	0	7	12	0	0	0	1
0	0	0	0	1	0	37	1	2	4
4	7	2	13	6	12	1	21	13	4
0	0	4	1	2	2	0	0	5	2
0	0	0	0	3	2	1	0	4	22

Confusion Matrix (Weizmann)

9	*0*	*1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
0	*8*	*0*	*0*	*0*	*0*	*0*	*0*	*1*	*2*
0	*0*	*6*	*0*	*0*	*0*	*1*	*0*	*1*	*0*
0	*0*	*0*	*9*	*0*	*1*	*1*	*0*	*0*	*0*
0	*0*	*0*	*0*	*9*	*0*	*1*	*0*	*0*	*0*
0	*0*	*0*	*0*	*0*	*8*	*0*	*1*	*0*	*0*
0	*0*	*2*	*0*	*0*	*0*	*5*	*0*	*0*	*0*
0	*0*	*0*	*0*	*0*	*0*	*1*	*8*	*0*	*0*
0	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*7*	*0*
0	*1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*7*

Improvements

- Why were the results for the Weizmann dataset (84.4%) so much better than the Boyer dataset (48.2%)?
  - Some of the actions in the Boyer set look the similar (e.g. the left and right woots), especially to a shape-only based descriptor
  - This can be fixed with a more advanced descriptor which incorporates shape, color, optical flow, and depth (currently not using any of those)
- Baseline results for my dataset: 91%
- Can we beat it?

Future Plans

- Improve the feature descriptor
  - Create 3 versions of it:
    - Color only
    - Depth only
    - Color and depth
- Create a website!