Week 7

Stephan Boyer
UCF REU

Quick Recap

- The ultimate goal is to be able to recognize a human action (out of 12 possible choices) as quickly as possible, i.e. in as few frames of a video sequence as possible.
- We know this is possible for certain actions with only a single frame.
- We want to investigate if depth information can be used to enhance the recognition.
- We are in the process of collecting a dataset of depth videos using a stereo camera called the Bumblebee.

Dataset Progress

- As you know, we ran into technical difficulties on Wednesday and were unable to complete the dataset.
- Problem: The Bumblebee camera does not work with FireWire ExpressCards.
- Problem: The Bumblebee requires a *powered* FireWire port, and most laptops does not provide this power.
- We have had limited recording success on other computers including two desktops and David's laptop.
- We have acquired a new computer with a FireWire port.
- So here's the problem:

Dataset Progress

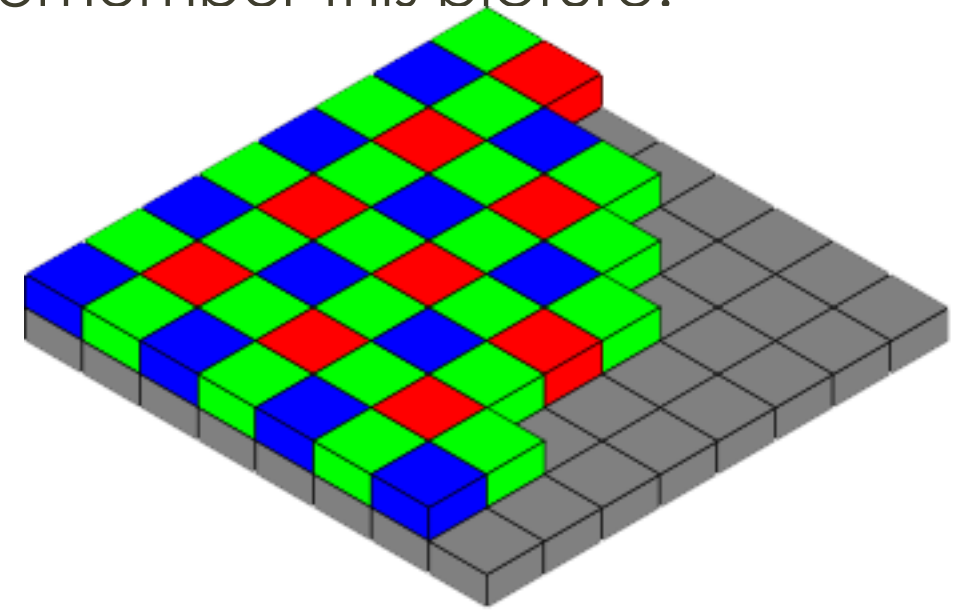
- The Bumblebee camera gives us raw mosaiced images at up to 48 Hz
- Let p be the number of pixels per frame (640x480)
- Let b be the number of bytes per pixel (3)
- Let f be the desired number of frames per second (30)
- Then $pbf = 640 \cdot 480 \cdot 3 \cdot 30 = 27.6$ Mb/s
- Most hard disks should easily be able to handle this, but only if it's written in *large chunks at a time*, for example copying large files etc.
- This is **not** what we're doing, and our frame rate suffers because of this.

Dataset Progress

- So why can't we write **compressed** video to disk instead of raw, uncompressed video?

Dataset Progress

- You remember this picture:



Dataset Progress

- Video compression artifacts will break the demosaicing algorithm! This means we have to write the videos **uncompressed**.
- Or do we?
- It is possible that a demosaicing algorithm could be executed on the GPU, in real-time.
- Then demosaicing could occur before writing the video to disk, which would enable us to use lossy compression techniques.

Dataset Progress

- Alternate possibility: Dump the raw data from the Bumblebee directly to disk without writing video file with OpenCV.
- Advantages:
  - Less data to be written (no video container headers etc.)
  - More flexibility regarding writing process (How much do we write at a time? Do we write after each frame, or after each 10 frames, …?)
- Disadvantages:
  - More processing later to convert to video format
- This is probably the best option.

Dataset Progress

I wrote a demosaicing algorithm that runs on the CPU.

Here's the original, raw video from the camera.

Why is it blue?
Why is it fast?

Dataset Progress

This is the result of my demosaicing algorithm, which uses bilinear interpolation.

New Problem

- Assume we have a good dataset.
- How do we segment out the person from his background?
- Techniques:
  - Optical flow (moving pixels are likely to be part of the person)
  - Similarly, histeresis (pixels that are static for many frames are likely to be part of the person)
  - Location-awareness (pixels near the center of the frame are more likely to be part of the person)
  - *Our contribution:* **Depth** (pixels near the camera are likely to be part of the person)

Segmentation Test

- For every pixel:
  - Let $\sigma = \alpha \cdot p_z + \beta \cdot (p_x - c_x)^2 + \gamma \cdot (p_y - c_y)^2$

    where α, β, γ are constants

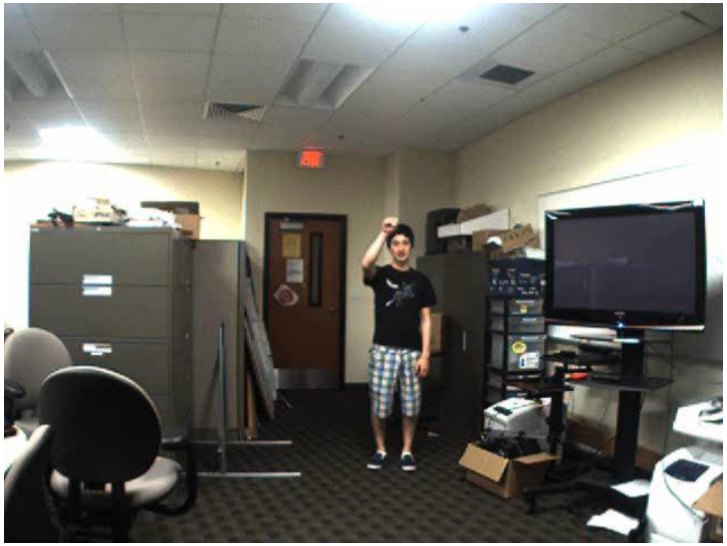
    c_x, c_y is the center of the frame

    p_x, p_y is the pixel location

    p_z is the pixel depth
  - If $\sigma < \varepsilon$ for some threshold ε, then the pixel is masked out.
  - The remaining pixels should be near the person.

Initial Results

Original Sequence



Segmentation

Other Activity

- Investigating the use of support vector machines as action classifiers (through the use of *libsvm*)
- Performing hard disk write speed tests to find the best way to write large amounts of data to disk in real time
- Investigating potential for refactoring dataset activities to meet the needs of Miko (one of Dr. LaViola's grad. students)

Conclusion

- New ideas (and hardware) for data capture should advance dataset progress.
- Segmenting the foreground person from the background is not a big deal.
- Initial depth estimates from stereo are somewhat disappointing – full of noise and inaccuracies (could be a calibration problem).