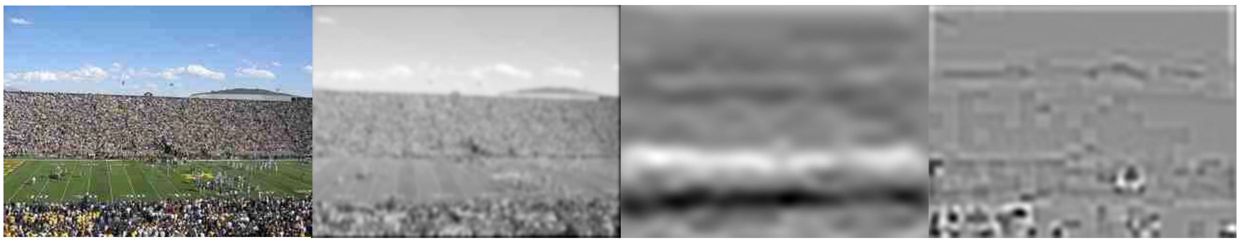
Basel Alghanem
16-720
Homework 2

1.1) There are four types of filters. The first is simply a gaussian filter. The smaller ones do less smoothing; more smoothing on the right. The second is a Laplacian of gaussian, which estimates the 2^{nd} derivative (good for edge detection). The third and fourth take derivatives in x and y. The smaller ones pick up sharper edges better, whereas the rightmost will do slower or noisier gradients better.

1.2) The Lab colorspace uses 3 channels. One is the lightness of the image, and the other two describe color on two spectrums: red-green and blue-yellow. One reason to use it is that it tries to replicate human vision (which is one way to get computer vision to work better), including by trying to maintain perceptual uniformity (size of change in values are proportional to size of change in importance) and having a definition of lightness that's similar to human understanding of lightness.



The selected image (of Michigan Stadium prior to its 2010 renovation) and three filter responses:

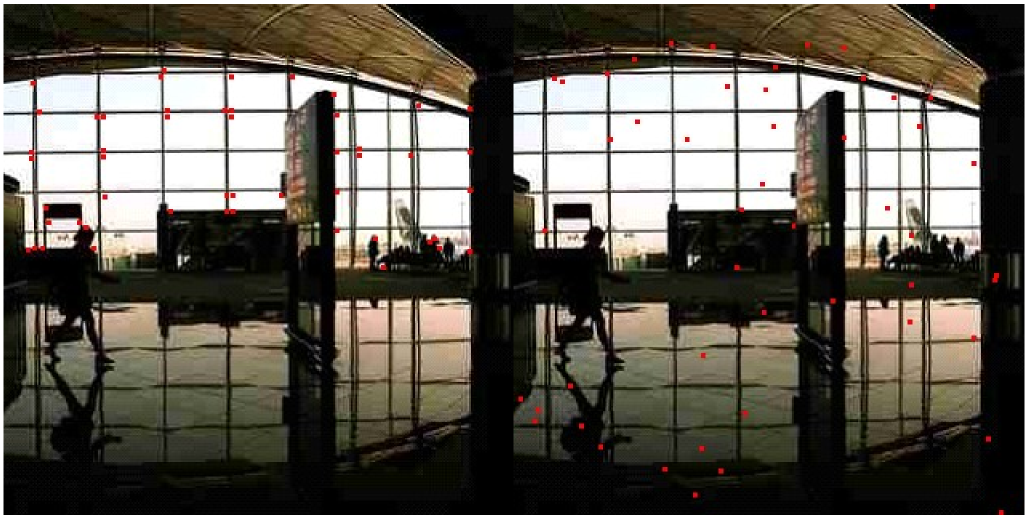
1. The fourth largest gaussian filter on the lightness channel
2. The second largest y-derivative filter on the a (green-red) channel
3. The third largest Laplacian of gaussian filter on the b (blue-yellow) channel

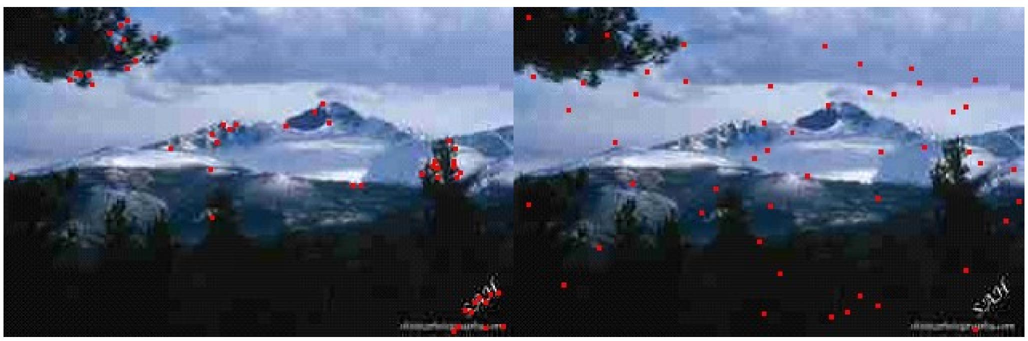
The jpg artifacts can have a big impact on the image. This is especially seen in the third selected filter response above, where the edges of the jpg blocks show prominently along with responses of the actual image (like the yellow block M at the center of the field).

1.3)

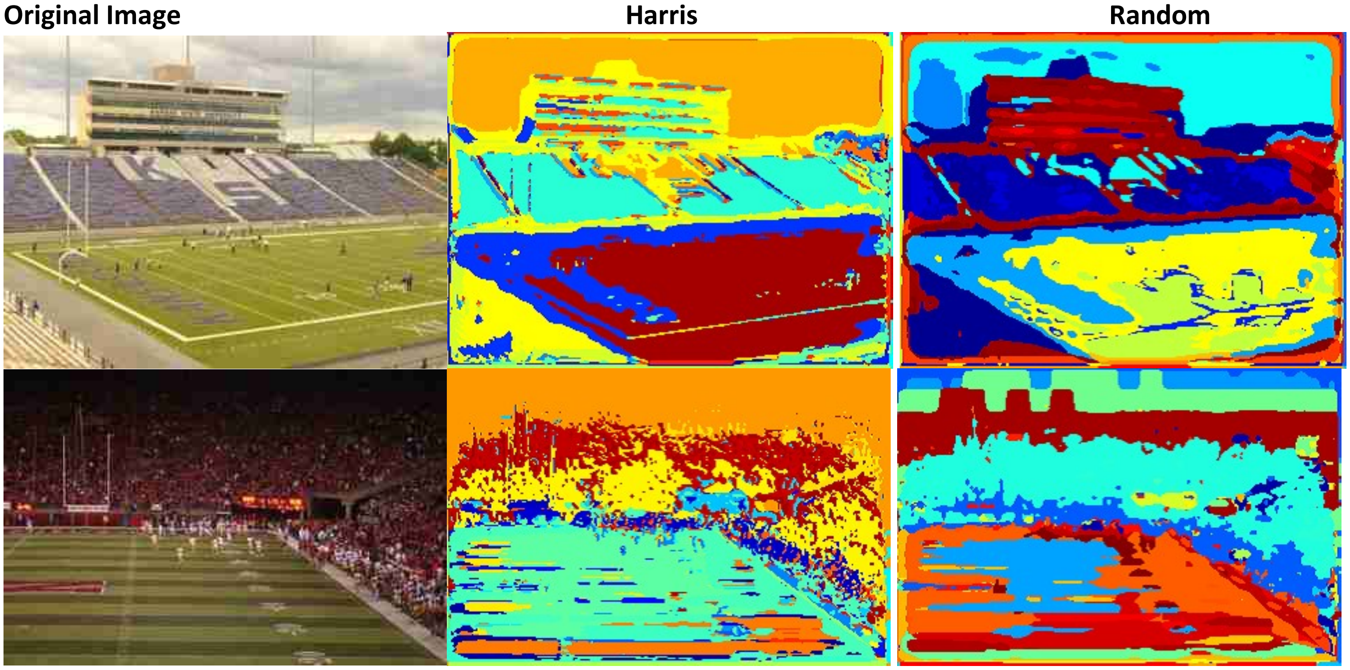
Basel Alghanem
16-720
Homework 2

1.3) cont.





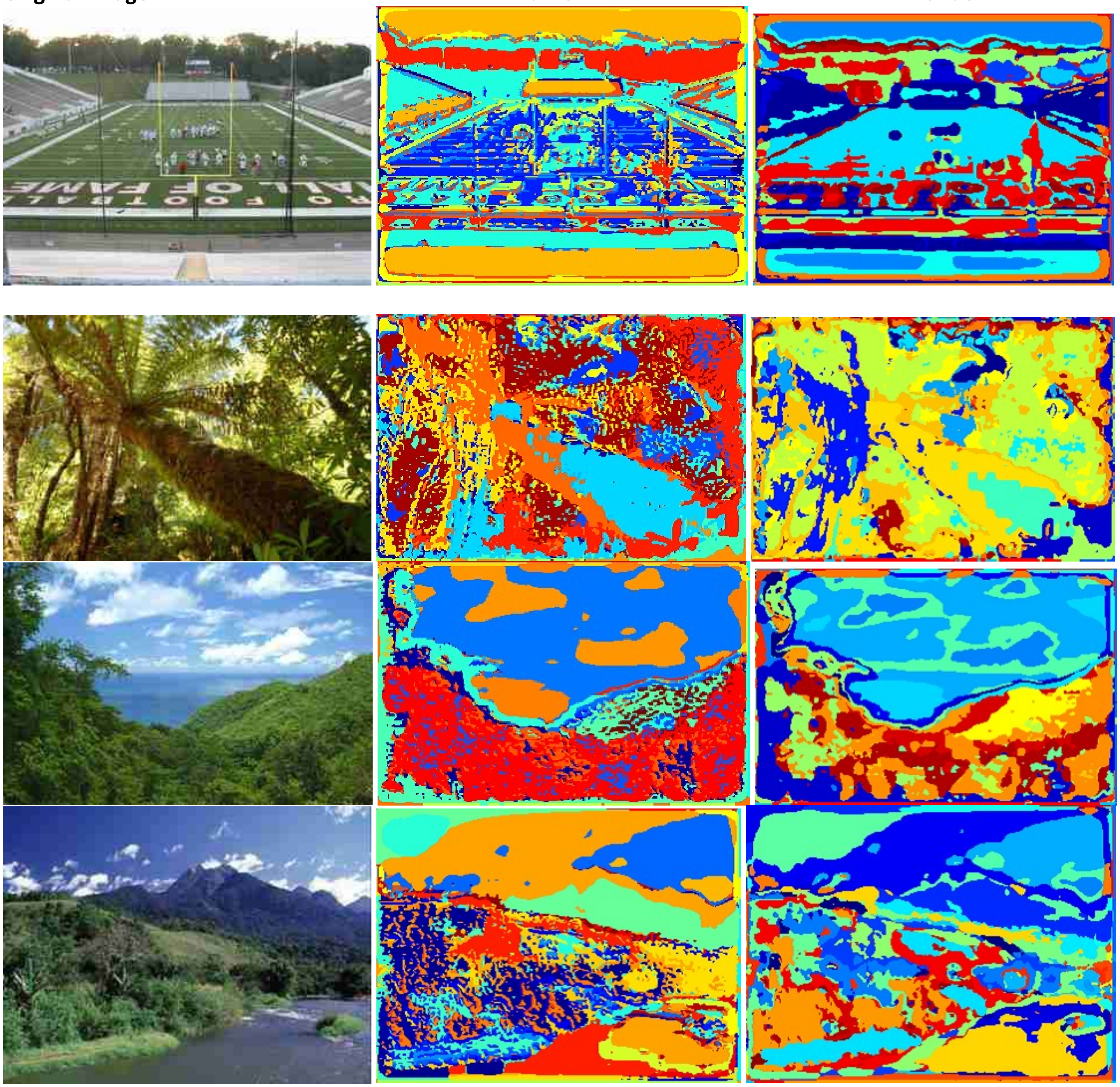
2.1) **Original Image** **Harris** **Random**



Basel Alghanem
16-720
Homework 2

2.1) cont.

Original Image **Harris** **Random**



The visual words are capturing meaning. In both classes, they can separate sky from other things (seats, field, mountain, foliage). It's hard to tell how well, because it's inconsistent. In image 5, the Harris dictionary separated sky from clouds very well, whereas the Random dictionary separated sky and clouds in several levels. The big difference between the dictionaries is that Random dictionary has larger patches, whereas some labels from the Harris dictionary look more like Jackson Pollock paintings. This makes me think the Random dictionary is working better.

Basel Alghanem
16-720
Homework 2

3.1) Just a note, I found and used an alternative pdist2 (calling it pdist3).

3.2)

Harris – Euclidean	Accuracy: 41% Confusion: $\begin{bmatrix} 7 & 5 & 2 & 1 & 1 & 2 & 1 & 4 \\ 2 & 11 & 5 & 2 & 2 & 1 & 1 & 0 \\ 3 & 0 & 10 & 2 & 2 & 1 & 1 & 0 \\ 3 & 0 & 1 & 5 & 1 & 5 & 9 & 3 \\ 0 & 3 & 0 & 0 & 9 & 1 & 2 & 0 \\ 1 & 1 & 0 & 3 & 0 & 5 & 0 & 0 \\ 1 & 0 & 2 & 4 & 5 & 4 & 5 & 0 \\ 3 & 0 & 0 & 3 & 0 & 1 & 1 & 13 \end{bmatrix}$
Random – Euclidean	Accuracy: 41% Confusion: $\begin{bmatrix} 11 & 6 & 6 & 2 & 0 & 0 & 3 & 2 \\ 1 & 11 & 3 & 1 & 4 & 3 & 4 & 0 \\ 3 & 1 & 9 & 1 & 1 & 1 & 3 & 0 \\ 2 & 0 & 1 & 5 & 3 & 4 & 3 & 2 \\ 0 & 1 & 1 & 2 & 8 & 2 & 0 & 0 \\ 0 & 1 & 0 & 5 & 1 & 4 & 0 & 1 \\ 1 & 0 & 0 & 3 & 3 & 2 & 3 & 1 \\ 2 & 0 & 0 & 1 & 0 & 4 & 4 & 14 \end{bmatrix}$
Harris – ChiSq	Accuracy: 46% Confusion: $\begin{bmatrix} 7 & 5 & 2 & 1 & 0 & 2 & 0 & 6 \\ 2 & 9 & 3 & 2 & 2 & 2 & 0 & 0 \\ 6 & 2 & 12 & 3 & 2 & 1 & 2 & 0 \\ 1 & 0 & 0 & 8 & 0 & 3 & 4 & 2 \\ 0 & 2 & 1 & 0 & 12 & 2 & 5 & 0 \\ 1 & 1 & 0 & 2 & 0 & 5 & 0 & 0 \\ 0 & 1 & 2 & 3 & 4 & 3 & 8 & 0 \\ 3 & 0 & 0 & 1 & 0 & 2 & 1 & 12 \end{bmatrix}$
Random – ChiSq	Accuracy: 48% Confusion: $\begin{bmatrix} 14 & 2 & 4 & 1 & 0 & 1 & 3 & 2 \\ 1 & 13 & 4 & 1 & 3 & 3 & 2 & 0 \\ 3 & 2 & 10 & 3 & 2 & 3 & 3 & 1 \\ 0 & 0 & 0 & 6 & 2 & 1 & 2 & 2 \\ 0 & 2 & 2 & 2 & 7 & 0 & 2 & 0 \\ 0 & 1 & 0 & 5 & 1 & 7 & 0 & 0 \\ 0 & 0 & 0 & 2 & 5 & 2 & 5 & 1 \\ 2 & 0 & 0 & 0 & 0 & 3 & 3 & 14 \end{bmatrix}$

The random dictionary did a little better with one of the two distance measurements, and they were the same for the other. This is not surprising, since the classification was based on every pixel, but the Harris dictionary only learned what different corners looked like, not what all kinds of pixels looked like.

The Chi Squared metric did much better than the Euclidean metric. The difference between the two is that the Chi Squared metric upweighs uncommon words and downweighs common words. I can rationalize why this would perform better — if an uncommon word appears at all often, that fact should be taken strongly into account — but I would not have necessarily expected it if I didn’t see the results first.