

Figure 3: The provided multi-scale filter bank

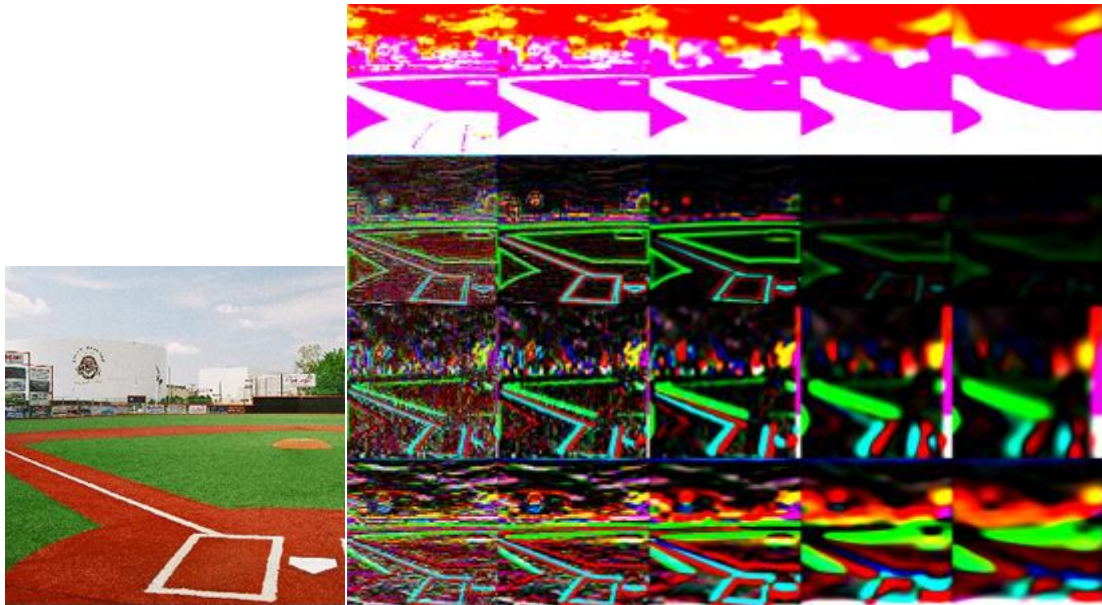
#### 1. FilterResponses

1<sup>st</sup> row filters are Gaussian filters, 2<sup>ed</sup> filters are Gaussian Laplacian filters. However, according to the filter color, the direction of Gaussian filters and G-L filters is opposite, which means if the 1<sup>st</sup> row filters are concave filters, the other one will be convex. Meanwhile, both filters don't have horizontal direction.

3<sup>rd</sup> row filters and 4<sup>th</sup> row filters are both Derivative of Gaussian (DoG) filters. The 3<sup>rd</sup> row filters are DoG filters on x-derivative direction, while 4<sup>th</sup> row filters are on y-derivative direction. DoG filters not only have vertical direction but also horizontal direction.

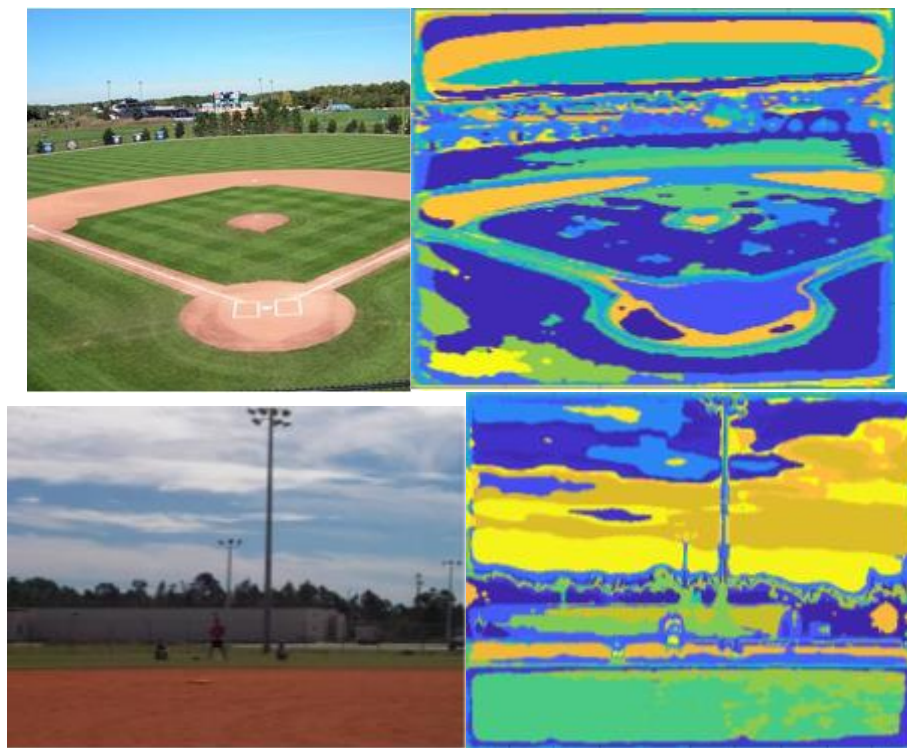
From left to right, the size of the filters increases.

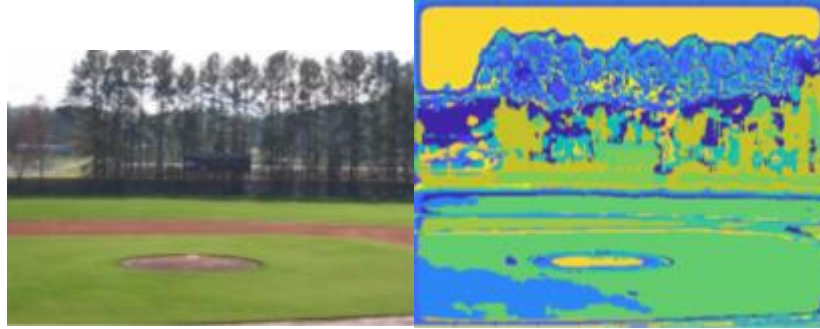
#### 2. Extracting FilterResponses



The left figure is the original image, and the right one shows its 20 filter responses in four different filters as well as five different scale.

### 3. Visual words





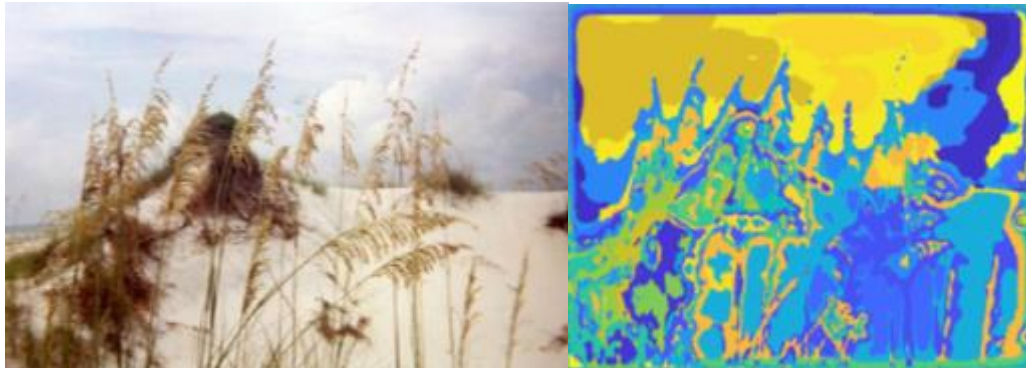
Three pairs of figures show original RGB images and their corresponding visual words. As showing in the image, the flat, smooth, constant parts of one scene are almost represented as one color which means this region could be represented by one word.

#### 4. Confusion matrix

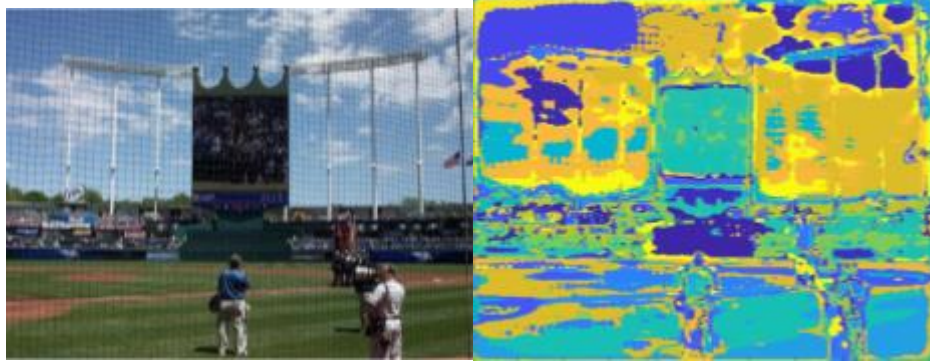
12	0	0	1	1	2	3	1
0	10	1	1	0	0	0	8
2	0	9	1	1	2	0	5
0	2	0	11	0	0	1	6
4	1	0	0	12	2	1	0
2	0	0	1	7	8	2	0
1	0	0	2	0	5	12	0
0	0	0	7	0	0	1	12

Accuracy=53.75%

#### 5. Find failed cases



Right image belongs to desert scene, but the prediction result is kitchen. Compared to other training, testing images, this desert was covered by several reeds which separate the desert into many small pieces. The visual words of it should have represented it as a smooth and flat region, but in this case, it will be harder to match this with most desert scene.



Left figure belongs to baseball field, but the prediction result is windmill. Other successful baseball field images are always separated into three parts, including green field, auditorium, and broad sky above them. To some extent, they are easier to be matched for this obvious character. However, this baseball field has a tall, big screen in the center of the scene, which is similar with windmill, where a tall building stands on the ground and is surrounded by blue sky.