Author: Francisco Vicente Carrasco, email:franciscovicencar@gmail.com

Object Bank feature description

I will try to explain the objectbank feature and how they obtain the feature. This is the explanation for the feature vector that corresponds to 1 object, the total feature vector is composed by the vector obtained for everyobject.

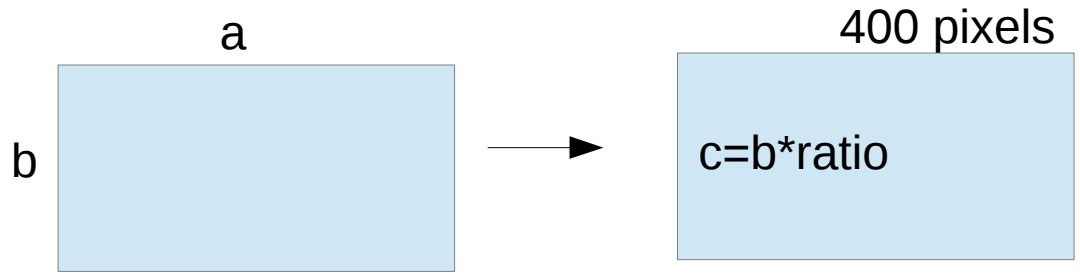
1 Object = 252 dimensions
Total descriptor= num_objects*252.

First step(Resizing the original image):

The image is resized using the following process. They get the image dimensions(i.e. (a,b)). The ratio for scaling is calculated, using the following:

Ratio=400/min(a,b);

Therefore, the smaller axe of the image is converted to 400 pixels. This example illustrates that:



For example: min(a,b)=a

Ratio=400/min(a,b)=400/a

Second step(Getting HOG features at different scales):

After this reescaling, HOG features are obtained using different scales of the image. Although, they obtain HOG features for more scales, they only use six of these scales. These are the ratios used for resizing the image(previously resized in the previous step)

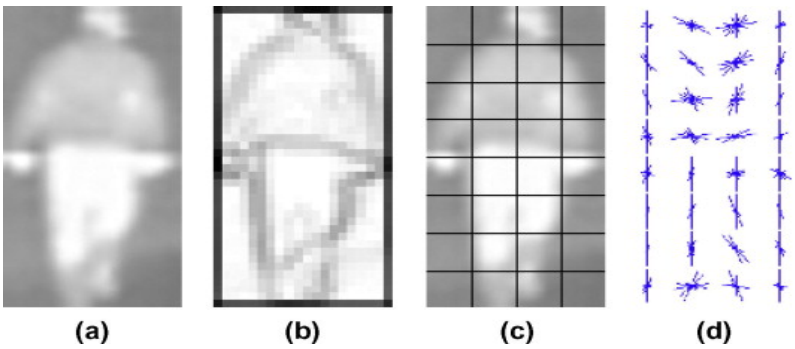
Ratios:
1(image obtained from the previous step)
0.707
0.5
0.3535
0.25
0.17677
0.128

As you can see, the image with the bigger resolution is the first one.

Author: Francisco Vicente Carrasco, email:franciscovicencar@gmail.com

After resizing the images, they calculate the HOG features for every image. These HOG features are used to obtain the response for every object.

Example of HOG features are calculated for one image:



Third step(Getting the response for the object):

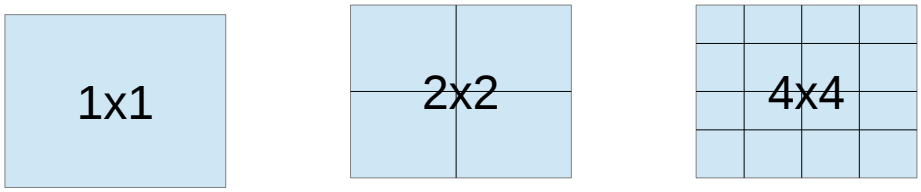
After getting the HOG features, we apply a object specific filter to these features. Each root filter, has two different components. Each of these components works on a different scale. As a result, we have 12 different detection scales because we obtained had 6 different scales from the previous steps, and every filter works at 2 different scales. Consequently, 6*2=12.

These filter responses, are stored in a matrixes following the same distribution as the HOG feature in the image d of the previous figure.

Then, for the HOG obtained from each ratio we have two different filter responses. Namely, we have 12 HOG responses.

Fourth step(Getting the spatial piramids):

We have 3 different spatial pyramid levels:



These three spatial pyramid levels are applied to the 12 different responses to one object.
In order to select the value for each box, they select the maximum response of the filter for every box. For instance, for the second level, they split the filter response using a grid of 2x2; They pick the maximum response inside every box.

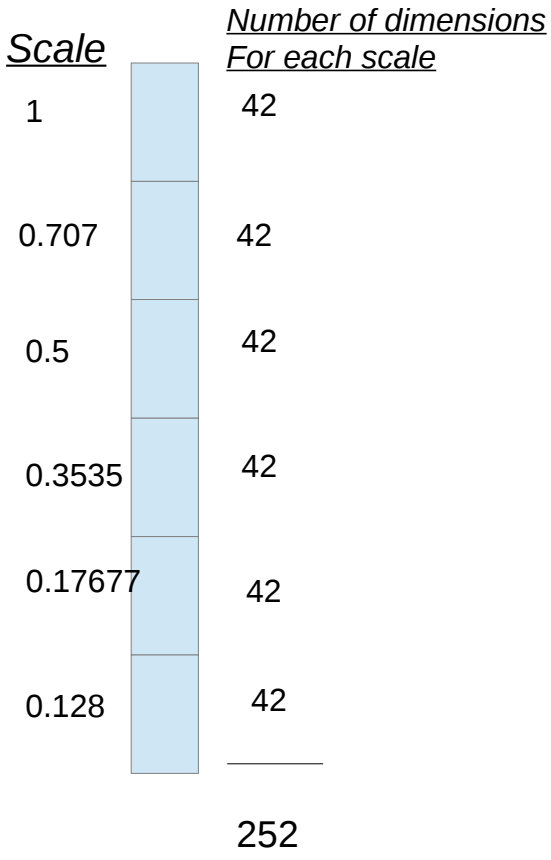
As you can see, we have 21(1 + 2*2 + 4*4) values for every one of the 12 filter responses for one object. Resulting in a total of 12*21= 252 dimensions vector for every object.

Author: Francisco Vicente Carrasco, email:franciscovicencar@gmail.com

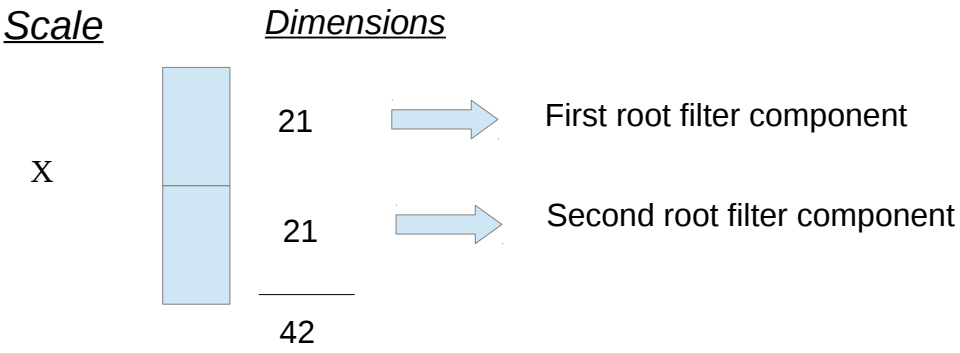
Fifth step(getting the feature vector for one object):

Now, I am going to describe the distribution of the feature vector for one object. We have a vector of 252 dimensions.

We start for the diferent scales(Remember that our original image, is the one obtained in the first step)



Each one of the chunk for every scale is divided in two pieces, because the used root filter in object bank has two different components that work at different scales. So these 42 dimensions for every scale are splited in two pieces of 21 dimensions.



Author: Francisco Vicente Carrasco, email:franciscovicencar@gmail.com

Finally, This is the distribution of the 21 dimensions.