We want to locate and track Vehicles in images of a video.

We will learn what features are the most useful for a supervised learning classifier that can help us detect whether something is a vehicle (or any object) or not.

For example: some characteristics in an image that might be useful to detect cars in an image are Color, Position within the image, shape, and Apparent Size and more!

We will learn about object Detection, and then learn how to classify them.

**Detection:**

Color and Gradient are good differentiators between objects. They are both features that we can use to help classify our image. In all applications, we will use a combination of features that lead to the best result.

So what features align with what characteristics in the objects we want to detect?

A Default images pixel values can give details on the color and shape of an object, while the histogram of pixel values gives us info on only the color. At the same time the gradients of pixel values gives us information about an objects shape.

**Features:**

Simplest Feature you can get from an image is the color. You can compare a known car image with the test region you want to check for that object. You can do this by subtracting pixel values or getting to correlation and threshold the values. This is called **Template Matching**.

The problem with template matching is that you need a Template. If you are thinking about tracking a vehicle from a video, its position relative to the camera changing as it speeds up and slows down would not match the template, meaning it won’t work well for us. (Same Orientation, size, and color)

We need to find a transformation that are robust to changes in appearance. An example of one transform is the **histogram of color values** in an image. It will match things in a region that share the same color distribution. By normalizing the distribution, we can also account for a variation in the objects size. It will also match the object if it is in slightly different aspects orientations can still be matched. Note that this is only matching based on color, which can still match unwanted regions.

Now let’s think about other color spaces. If we thing about HSV images. In some images, cars are usually saturated in color while the background is pale. Maybe Saturation values can help identify cars invariant of colors. Of course this isn’t always the case, but it’s good to try as much as you can.

You should get used to plotting your test images in different color spaces to try and determine which color spaces can help make vehicle detection color invariant.

There are many large datasets available of images used throughout the process of making autonomous vehicles. An example of one is <http://www.cvlibs.net/datasets/kitti/>

We can use some of the frames of video they have collected to see if we can locate clusters of colors that correspond to specific cars, trees, the sky, etc. This way we can get a better idea of what sections we can use to identify vehicles over the environment. This doesn’t need to be perfect but it will be one of many useful features that will be used to help us identify vehicles.

It may be confusing to look at the 3D plots of full images and try and gain information about the vehicles versus the environment, so you should also try to compare the 3D plots of vehicles versus the environment.

**Spatial Binning of Color**

Raw pixel values are quite useful to include in your feature vector in searching for cars.

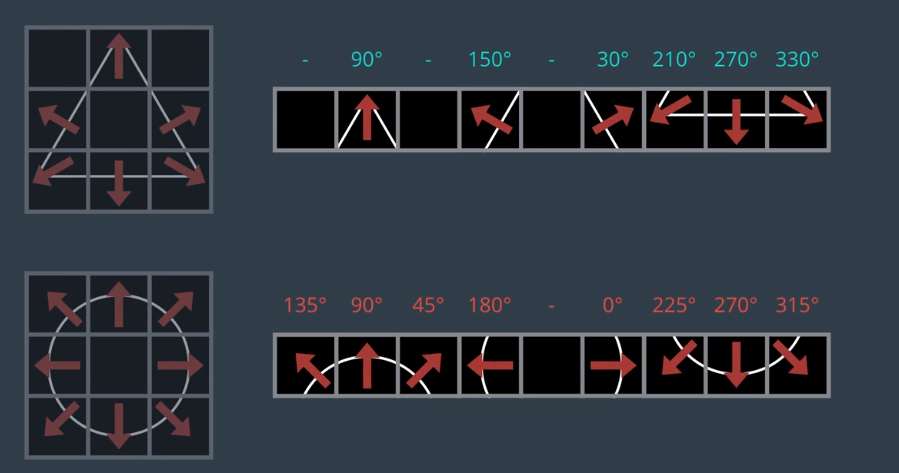


If you want to include raw pixel values, it can be cumbersome to include all 3 color channels of a full resolution image, so we can perform **spatial binning** on an image and still retain enough information to help in finding vehicles.

Spatial binning means to store information/pixel values into a combined bin in a way to organize it. Down sampling an image is an example of spatial binning. We can see from the image above that even going all the way down to 32x32 pixel resolution, the car is still clearly identifiable by eye, meaning relevant features are still preserved at this resolution.

**Gradient Features:**

When we want to gain information on a class object that can come in a variety of colors, structural queues like gradients and edges can give us a more robust representation. Actually the presence of gradients in specific direction around the center can give us information about its shape.



If we take a shapes gradient image, and separate it into different grid cells, and treat them as a flat 1D array, we can obtain a signature for that image. Different signatures correlate with different shapes. The signature for an image should have enough flexibility to account for small variations in orientation, size, etc. It should also be distinct so we can distinguish between shapes.

The above image of the shapes shows that the looking at the presence of gradients in specific directions around the center and lead to giving us a signature for that specific shape.

Using gradient values directly can lead to making the signature too sensitive.

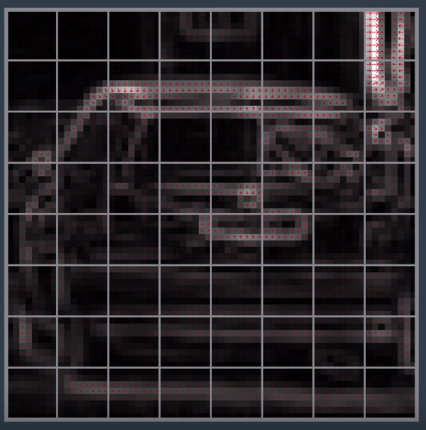
So to generalize the gradient and make the signature for each shape less sensitive, we use a technique called the **Histogram of Oriented Gradients (HOG).**

**Histogram of Oriented Gradients:**

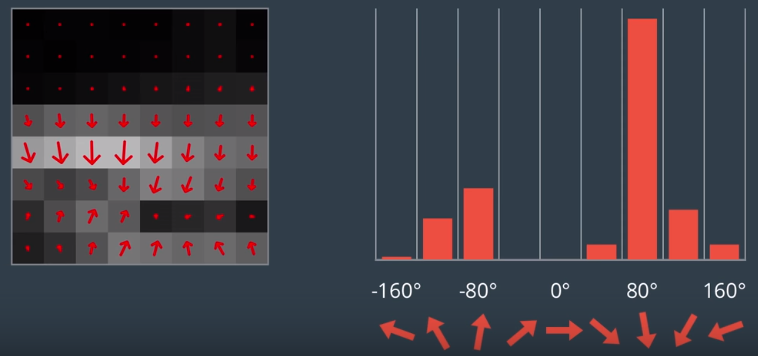
Getting the HOG of an image means to take the gradient of an image so that you can have the magnitude and direction (orientation) of the gradient. Separate your image into small cells. For each cell, you take its pixels gradients orientation and magnitude, and create a histogram of orientations for the whole cell. So in essence you are taking a histogram of the gradients orientations (Hence the name).

This gives us a histogram, meaning binned data, by the orientation like the image above where the magnitudes are summed.

Example:

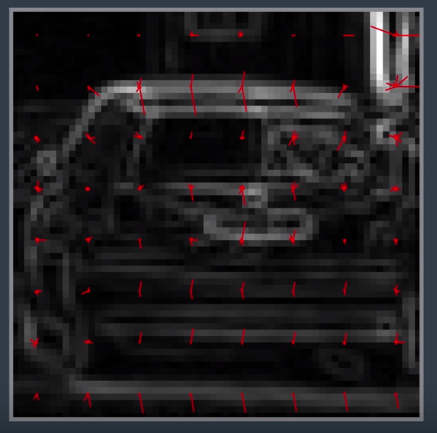
If we look at a 64x64 image of a car (1st image) and take its gradients magnitude and direction at each pixel, and then group them up into small cells of 8x8 each (2nd image). Then looking at each cell, we compute a histogram of the gradient directions (orientations) with each of the 64 pixels inside the 8x8 cell. The gradients samples at this small scale can be summed up into say 9 bins.



This histogram is not the sum of the number of pixels that fit into that orientation, but the sum of the magnitudes of those pixels.

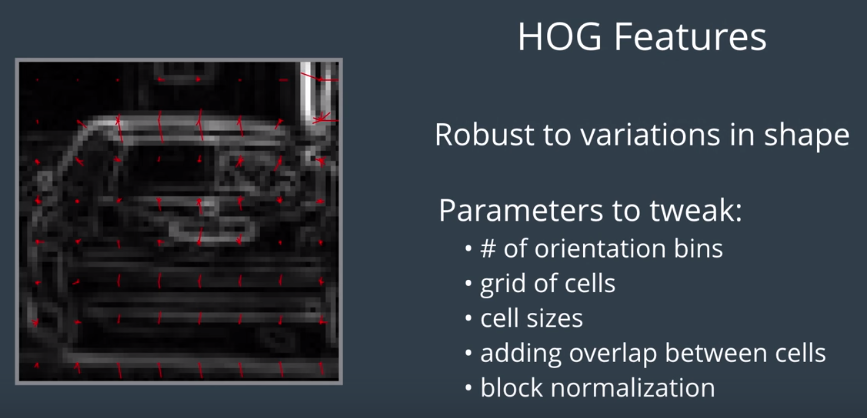


Each cell can be thought of an addition of the contributions from samples from each bin of the histogram to get a star of different lengths. The max length is the dominant gradient direction in this cell. Because the histogram is a sum of the magnitude values, the stronger gradient magnitudes contribute more weight to their orientation bin, so the effect of small random gradients due to noise is reduced.

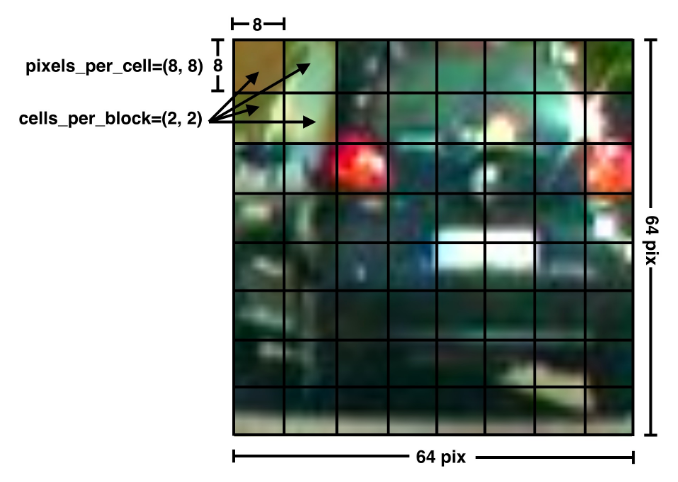


When you do this for all the cells, you can see a representation of the original structure emerge, which can become a key signature for a given shape.

This representation of the shape is called the Histogram of Orientated Gradients (HOG). It is then flattened into a vector and is used as a feature descriptor.



These parameters can be tweaked to change how accommodating/sensitive the feature is. In practice the intensity in a small block of cells is normalized to give a more unique but robust signature.



The scikit-image package has a build in function that extracts the Histogram of Oriented Gradients feature vector. The function can take the parameters shown in the image above, as well as another optional power law or “gamma” normalization scheme (set by the flag transform\_sqrt) that may help reduce the effects of shadows or other illumination variation, but will cause errors if your image has negative values.

So if you think about the number of elements in your HOG features vector for the whole image, instinct is to say that each cell has 9 bins, and there are 8x8 cells, so then the HOG features vector is 9x8x8= 576.

That is not true with block normalization, which calculates at each block, and the step across and down.

Meaning that in this image you can look and see that the blocks would be 4x4, but with normalization, the number of blocks becomes 7x7 due to the addition of one step down per block and one step across.  
Therefore, the total size of the HOG features vector is 7x7x2x2x9 = 1764. (2x2 cells per block)

This will serve as one of the features that you will use to train your SVM Classifier. This means that there will be a database used with images labeled as vehicles, and non-vehicles.

Images used from the project are comprised of images taken from: <http://www.gti.ssr.upm.es/data/Vehicle_database.html>

<http://www.cvlibs.net/datasets/kitti/>