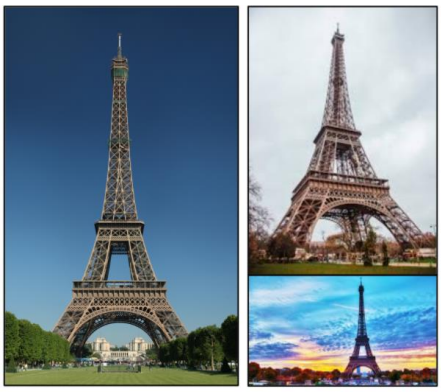
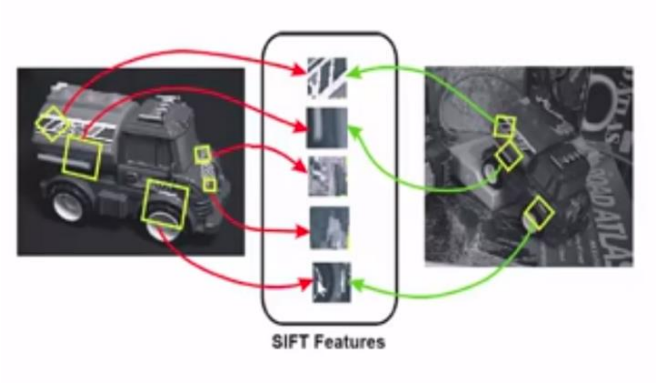
SIFT (Scale Invariant Feature Transform)





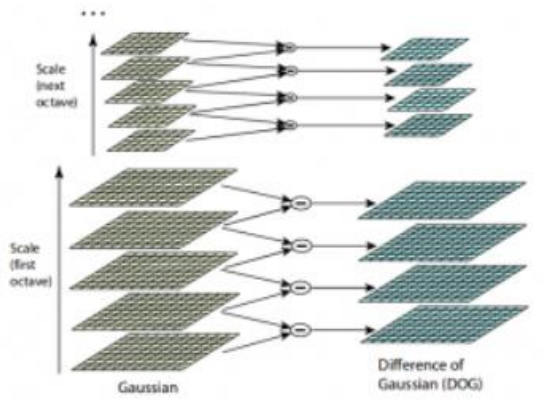
In the first set of images we see, Eiffel Tower present in everything. But the difference is the Tower is in different scale and orientation. SIFT is a Feature Detection and Description method, which identifies the key points in a image. These key points generated can be matched with another set of key points. This allows the computer to recognize the same object in different scales. In the second picture, we see the matching od key points in two different images.

We can also use the key points generated using SIFT as features for the image during model training. The major advantage of SIFT features, over edge features or hog features, is that they are not affected by the size or orientation of the image.

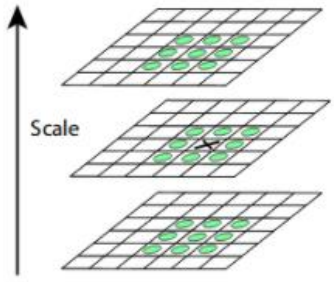
the entire process can be divided into 4 parts:

- **Constructing a Scale Space:** To make sure that features are scale-independent
- **Keypoint Localisation:** Identifying the suitable features or keypoints
- **Orientation Assignment:** Ensure the keypoints are rotation invariant
- **Keypoint Descriptor:** Assign a unique fingerprint to each keypoint

Constructing a Scale Space, this involves image pyramids. The image scaled is upto 4 octaves. In each Octave, 5 different scales are done. The Scaling involves Gaussian blurring, by altering the sigma values. The Gaussian blurring is used to reduce the noise. Just as Laplacian pyramid in Image pyramid, concept. A set of new images is got here. But that is done by the difference of Gaussian between the images of the same octave, and not between octaves. This produces an edge image.



Keypoint Localization. Here the new set of images is feed, to find the keypoints. Keypoints are found by local extrema. A pixel in a scale is compared with its neighbouring 8 pixels and 9 pixels each in the scales above and below. If the pixel is a local minima or maxima, it is a keypoint.



The keypoints can have low contrast and edges, these need to be removed. Taylor's series is done. The keypoints below a threshold value is discarded, next a second order Hessain matrix is done to discard the poor keypoints m such as edges or near edges. This method is just like the one in Harris Corner detection.

Orientation Assignment. Here a neighborhood around the keypoint is taken, and the magnitude and oreintaion is found.

$$\text{Magnitude} = \sqrt{[(G_x)^2+(G_y)^2]} = 16.64$$

$$\Phi = \text{atan}(Gy / Gx) = \text{atan}(1.55) = 57.17$$

The above is an example taken from a sample where the Gx = 9 and Gy = 14

After this histogram is plotted for magnitude and orientation. The highest peak is selected, also any peak above 80% is also selected. This stage can introduce new points, as number of peaks can be above 80%.

Now keypoint descriptor is created. A 16x16 neighbourhood around the keypoint is taken. It is devided into 16 sub-blocks of 4x4 size. For each sub-block, 8 bin orientation histogram is created. So a total of 128 bin values are available. It is represented as a vector to form keypoint descriptor.

Now a unique fingerprint is created. This can then be used for feature matching.