Explain How SURF is different from SIFT ?

Answer:-
The **scale-inavariant feature transform(SIFT)** an algorithm used to detect and describe local features in digital images. It locates certain key points and then furnishes them with quantitative information (so-called descriptors) which can for example be used for object recognition. The descriptors are supposed to be invariant against various transformations which might make images look different although they represent the same objects.

A **SIFT feature** is a selected image region (also called keypoint) with an associated descriptor. Keypoints are extracted by the SIFT detector and their descriptors are computed by the SIFT descriptor. It is also common to use independently the SIFT detector (i.e. computing the keypoints without descriptors) or the SIFT descriptor

SURF is the speed up version of SIFT. SURF goes a little further and approximates LoG with Box Filter. One big advantage of this approximation is that, convolution with box filter can be easily calculated with the help of integral images. And it can be done in parallel for different scales. Also, **the SURF rely on determinant of Hessian matrix for both scale and location.** For orientation assignment, SURF uses wavelet responses in horizontal and vertical direction. Adequate guassian weights are also applied to it.The dominant orientation is estimated by calculating the sum of all responses within a sliding orientation window of angle 60 degrees. wavelet response can be found out using integral images very easily at any scale. SURF provides such a functionality called Upright-SURF or U-SURF. It improves speed and is robust . OpenCV supports both, depending upon the flag,upright. If it is 0, orientation is calculated. If it is 1, orientation is not calculated and it is faster.

2. Briefly explain the main principles of FLANN matching and RANSAC (5 sentences)

Answer:-

FLANN matching :-

FLANN (Fast Library for Approximate Nearest Neighbors) is an image matching algorithm for fast approximate nearest neighbor searches in high dimensional spaces. **These methods project the high-dimensional features to a lower-dimensional space and then generate the compact binary codes.**

It contains a collection of algorithms optimized for fast nearest neighbour search in large datasets and for high dimensional features. It works faster than BFMatcher for large datasets. However, it's not going to find the best possible matches. Instead, it's just going to find good matching candidates.

FLANN builds an efficient data structure (KD-Tree) that will be used to search for an approximate neighbour, while BFMaatcher does an exhaustive search and is guaranteed to find the best neighbour. The real benefit of FLANN is seen with large data sets.

RANSAC:-

Random sample consensus or RANSAC is an iterative method for estimating a mathematical model from a data set that contains outliers. The RANSAC algorithm works by identifying the outliers in a data set and estimating the desired model using data that does not contain outliers. RANSAC is accomplished with the following steps

1. Randomly selecting a subset of the data set
2. Fitting a model to the selected subset

3. Determining the number of outliers
4. Repeating steps 1-3 for a prescribed number of iterations

RANSAC uses the voting scheme to find the optimal fitting result. Data elements in the dataset are used to vote for one or multiple models. The implementation of this voting scheme is based on two assumptions: that the noisy features will not vote consistently for any single model (few outliers) and there are enough features to agree on a good model (few missing data).

An advantage of RANSAC is its ability to do robust estimation of the model parameters i.e. it can estimate the parameters with a high degree of accuracy even when a significant number of outliers are present in the data set. A disadvantage of RANSAC is that there is no upper bound on the time it takes to compute these parameters . When the number of iterations computed is limited the solution obtained may not be optimal, and it may not even be one that fits the data in a good way. In this way RANSAC offers a trade-off by computing a greater number of iterations the probability of a reasonable model being produced is increased. Moreover, RANSAC is not always able to find the optimal set even for moderately contaminated sets and it usually performs badly when the number of inliers is less than 50%.

3. Explain the procedure and your approach and observations ? -- Assignment 2-b

Answer:-

The concept of “Bag of Visual Words” is taken from the related “Bag of Word” concept of Natural Language Processing.
In the bag of word model, the text is represented with the frequency of its word without taking into account the order of the words (hence the name ‘bag’).

Bag of Visual Words:-

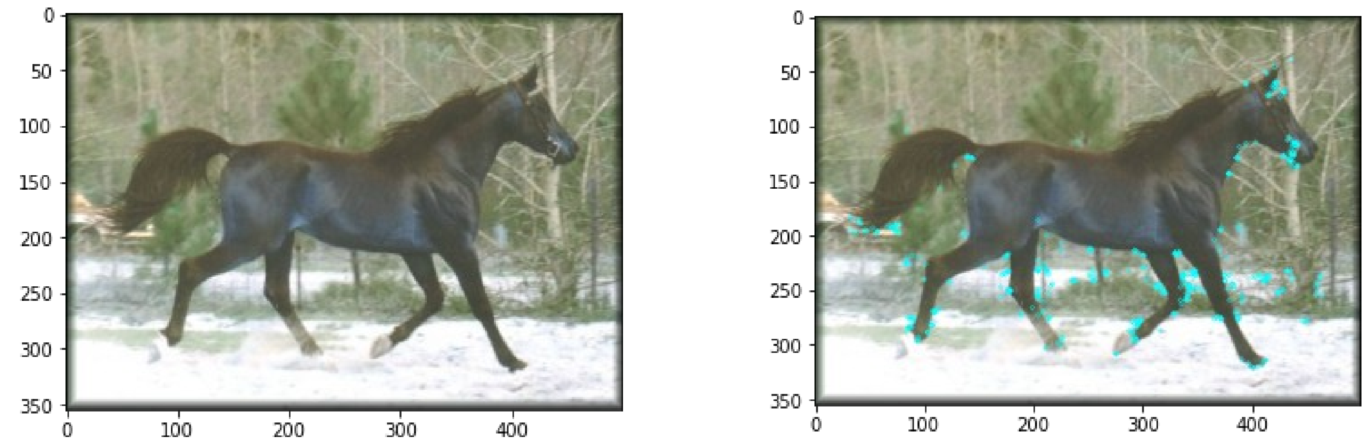
In Computer Vision, the same concept is used in the bag of visual words. Here instead of taking the word from the text, image patches and their feature vectors are extracted from the image into a bag. Features vector is nothing but a unique pattern that we can find in an image.

What is the Feature? :-

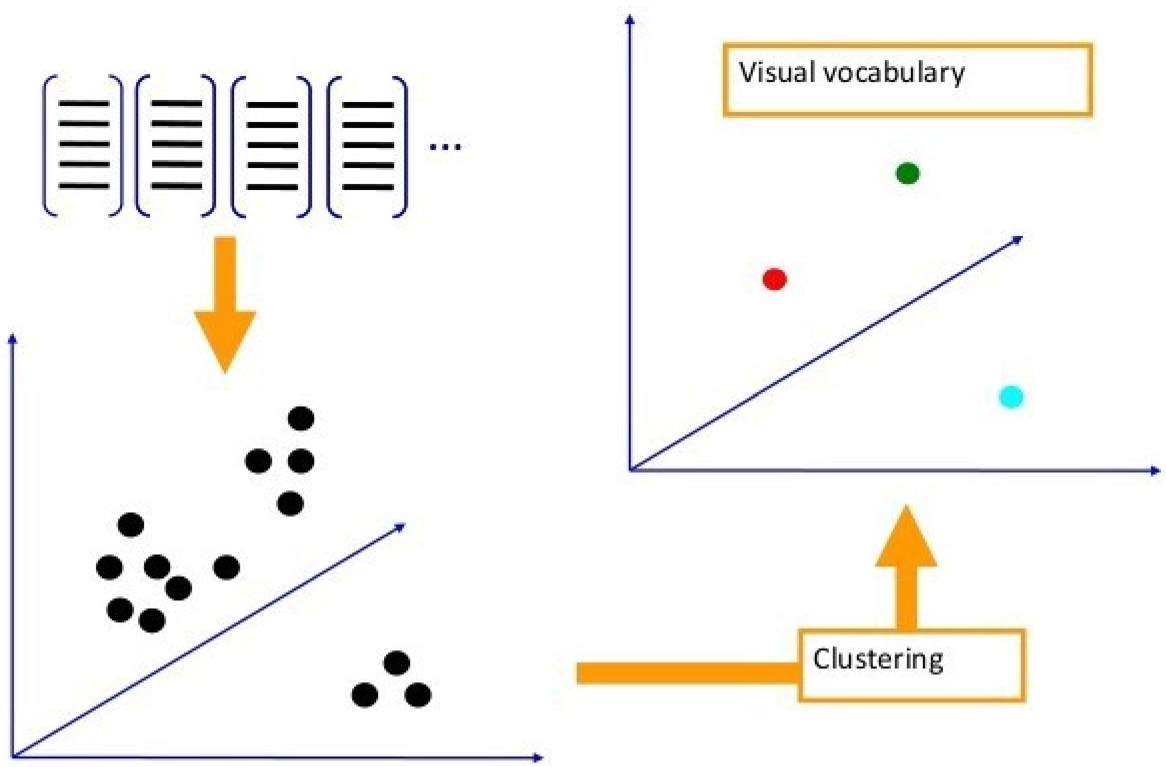
Basically, the **feature of the image consists of keypoints and descriptors. Keypoints are the unique points in an image, and even if the image is rotated, shrink, or expand, its keypoints will always be the same.** And descriptor is nothing but the description of the keypoint. The main task of a keypoint descriptor is to describe an interesting patch(keypoint)in an image.

Step by Step Bag Of Visual Words:-

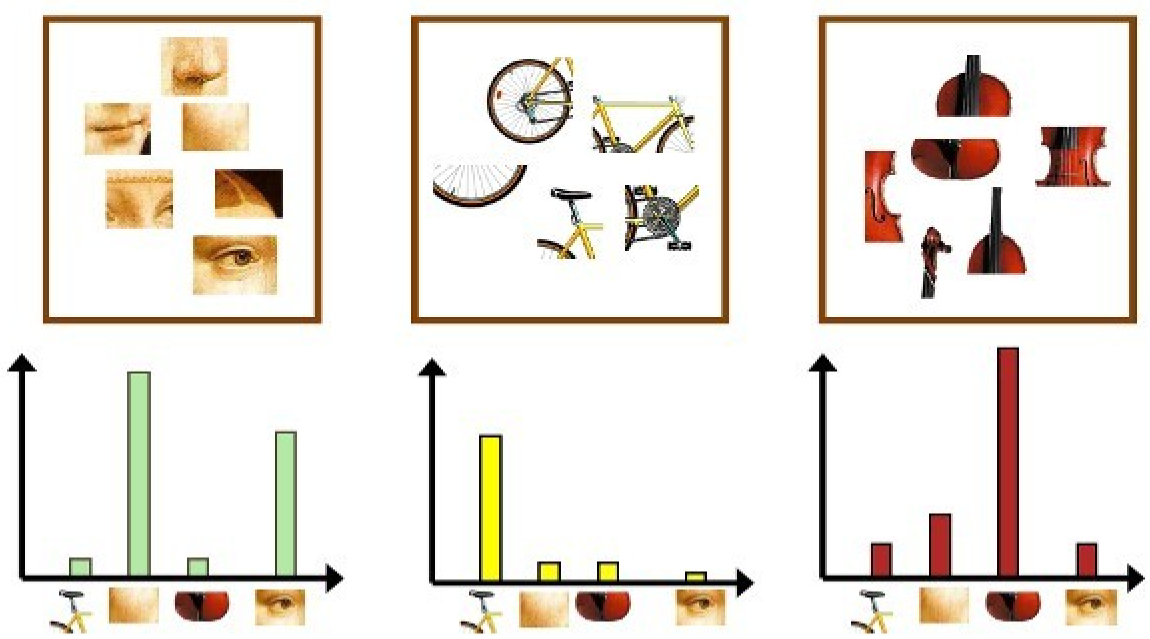
1. We extract local features from several images using SIFT/SURF/ORB.



2.Quantize the feature space. **Make this operation through clustering algorithms** such as **K-means**. The center points, that we get from the clustering algorithm, are our visual words.



3. Extract local features and compare these features with visual words to **create histograms for each image both for the test and train dataset**. It is done by first applying the keypoint detector or feature extractor and descriptor to every training image, and then matching every keypoint.



4. In this way, an image can be represented by a histogram of codewords. The histograms of the training images can then be used to learn a classification model. Here I am using **SVM as a classification model.**

As we can see in my code section , i am able to achieve an accuracy of 86% with this classical technique of image classification with bag of visual words model.