Ques 2a.

SURF v SIFT :

- SURF uses integral images for speeding up the calculations.
- SIFT builds image pyramids, filtering each layer with Gaussians of increasing sigma values and taking the difference. On the other hand, SURF creates a “stack” without 2:1 down sampling for higher levels in the pyramid resulting in images of the same resolution. Due to the use of integral images, SURF filters the stack using a box filter approximation of second-order Gaussian partial derivatives, since integral images allow the computation of rectangular box filters in near constant time.
- SURF uses Hessian Determinant for finding Keypoints.
- SURF calculates 64 dimensional feature vector whereas SIFT calculates 128 dimensional feature vector.

RANSAC:

RANSAC algorithm is a learning technique to estimate parameters of a model by random sampling of observed data. Given a dataset whose data elements contain both inliers and outliers, RANSAC uses the voting scheme to find the optimal fitting result.
It achieves it goal by:

- Selecting a random subset of the original data and fitting a model to the set of these hypothetical inliers.
- All other data are then tested against this model. Those points that fit the estimated model well, are considered part of the consensus set. The estimated model is reasonably good if sufficiently many points have been classified as part of the consensus set.
- The procedure is repeated a fixed number of times, each time producing either a model which is rejected because too few points are part of the consensus set, or a refined model together with a corresponding consensus set size.

RANSAC can estimate the parameters with a high degree of accuracy even when a significant number of outliers are present in the data set but there is no upper bound on the time it takes to compute these parameters. RANSAC can estimate only one model for a particular data set.

FLANN:

FLANN is a library containing collection of algorithms optimized for fast neigbhor search in large datasets and for high dimensional features and a system for automatically choosing the best algorithm and optimum parameters depending on the dataset.

Ques 2b.

Image classification using bag of visual words approach:

Step 1:
Creating a set of descriptors vectors for each image. We have used SIFT algorithm of opencv for this purpose. SIFT gives us a 128 dimensional descriptor vector for each image.

Step 2:
Creating the feature space for all images. In this step we throw all of our descriptor vectors into a bag and perform clustering using Kmeans to pull out k groups of related features. The resulting cluster will be our set of features.

Step 3:
Computing the histogram of features for each image. This will give us a k-bin histogram for each image, representing the number of each feature contained in the image.

Step 4:
Fit a classifier. We have used SVM as a classifier.

Approach:
The approach for the solving the problem has been same as outlined in the aforementioned steps. We create 2 files : learn.py and classify.py.
In learn.py step 2 (clustering) gives us a codebook which is then fed to the histogram to create feature bins. After step 3, we write and dump the histogram content and labels into a file as a visual vocabulary which is then passed to the svm for classification.

In classify.py, for a given image, descriptor vectors are obtained using SIFT and then histogram is compute for this image using the precomputed visual vocabulary. Then this is passed to the svm for classification.

Commands to use :
python q2b_learny.py -d path_to_image_folder/

python q2b_classify.py -c path_to_image_folder/codebook.file -m path_to_image_folder/trainingdata.svm.model image_you_want_to_classify

The path_to_image_folder should contain the training images already classified in different folders labelled by their class name.

The svm used in this program is taken from http://www.csie.ntu.edu.tw/~cjlin/libsvm by Chih-Chung Chang and Chih-Jen Lin.
The libsvm.py is taken from the libsvm packet by Chih-Chung Chang and Chih-Jen Lin.

Observation :

Although a very good approach to classifiy image based on the bag of words idea of classifying text data, this does not give good results when used on a wide variety and class of images as was seen when experimented with CIFAR-10 dataset. The accuracy for the bike vs horse classification was very good but with the CIFAR-10 dataset the accuracy dropped dramatically. Tweaking and improving the features and the kernels used in the SVM may help improve the accuracy.