International Institute of Information Technology, Bangalore

Assignment II
DS/NC/ESD 863 Machine Perception

Machine Learning

Group 7 - Team 14

Anoop Toffy - MT2016016
Athul Suresh - MT2016030

Instructor :
Prof. Dinesh Babu J

March 22, 2017

Contents

A Segmentation **2**
A.1 Procedure 2
A.2 Observations 3
A.3 Inference 8

B Panorama **9**
B.1 Observations 9
B.2 Inference 10
B.3 SURF vs SIFT 11
B.4 Principles of FLANN matching 12

C Bikes vs Horses Classification **13**
C.1 Procedure 13
C.2 Observation 15
C.3 General Inferences 16
C.3.1 Inference 1 16
C.3.2 Inference 2 16

List of Figures

6 Where segmentation fails 8
7 Panorama One 9
8 Panorama Two 10
9 Classification Bikes vs Horses 15

List of Tables

A Segmentation

- Pick 5 images from Berkeley Segmentation dataset
- Run Kmeans and Meanshift
- Show results along with the ground-truth

Answers

A.1 Procedure

- Convert the image into HSV colorspace.
- Apply Median blur with a big kernel to remove spurious noise and suppress small details.
- Convert the image into Mx3 feature matrix where M=height*width. The feature matrix contains M data-points(pixels) where each data-point represents (R,G,B) value of that pixel.
- Feed the feature matrix into K-Means/Mean Shift to get the cluster centers.
- Reconstruct the image from feature matrix, replacing each pixel value with value of the cluster centroid it belongs to.
- Use Laplacian edge detection to find the segments.
- Superimpose the detected segments on to the original image.

A.2 Observations

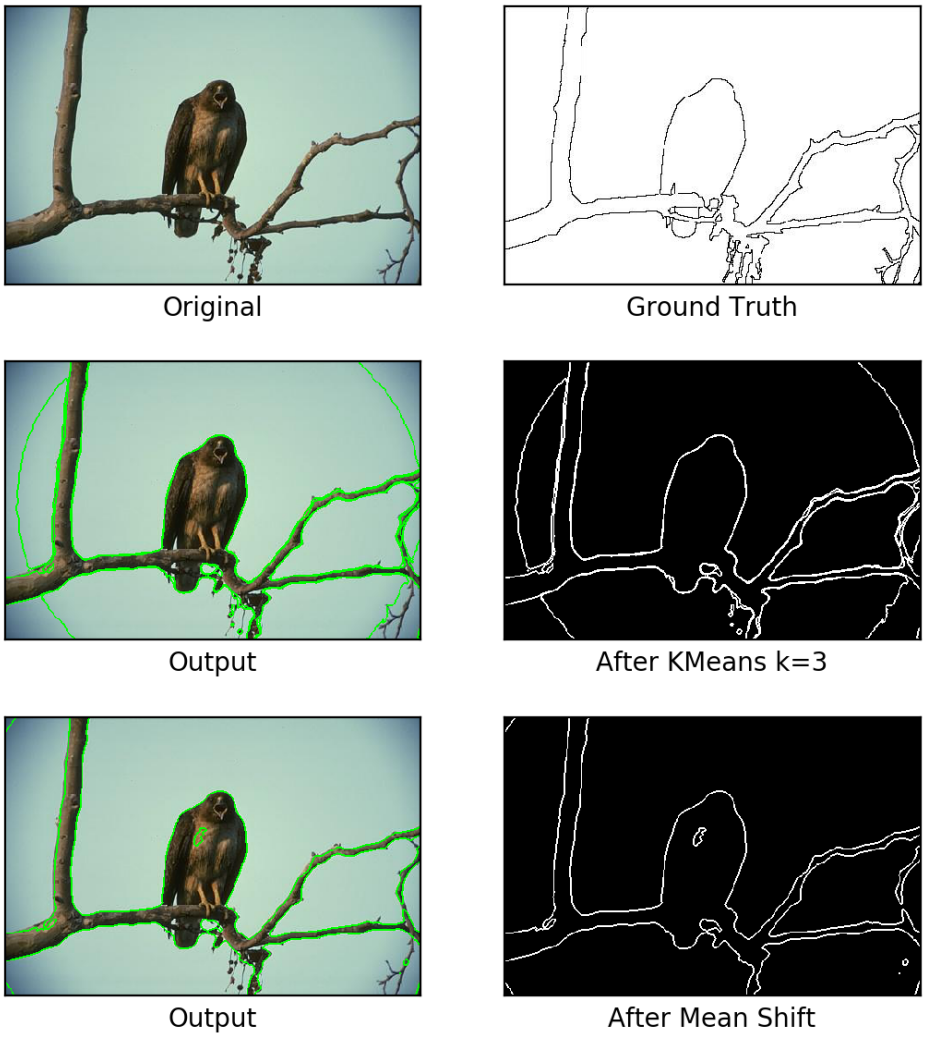

Figure 1

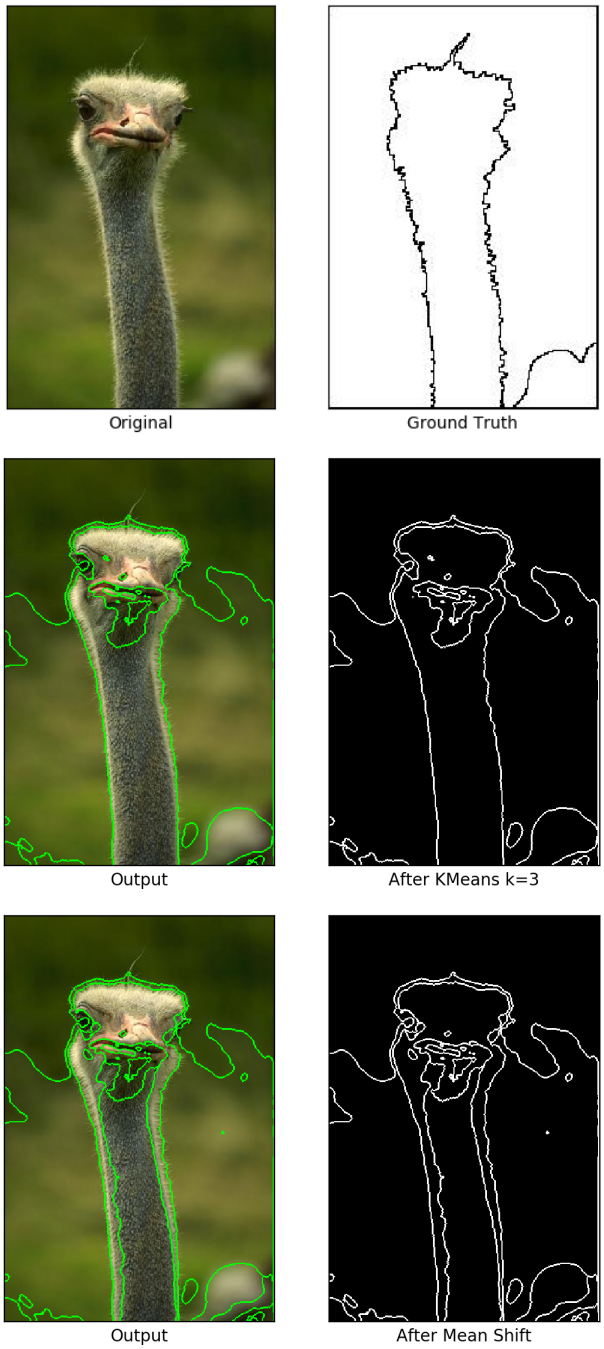
Original
Ground Truth
Output
After KMeans k=3
Output
After Mean Shift

Figure 2

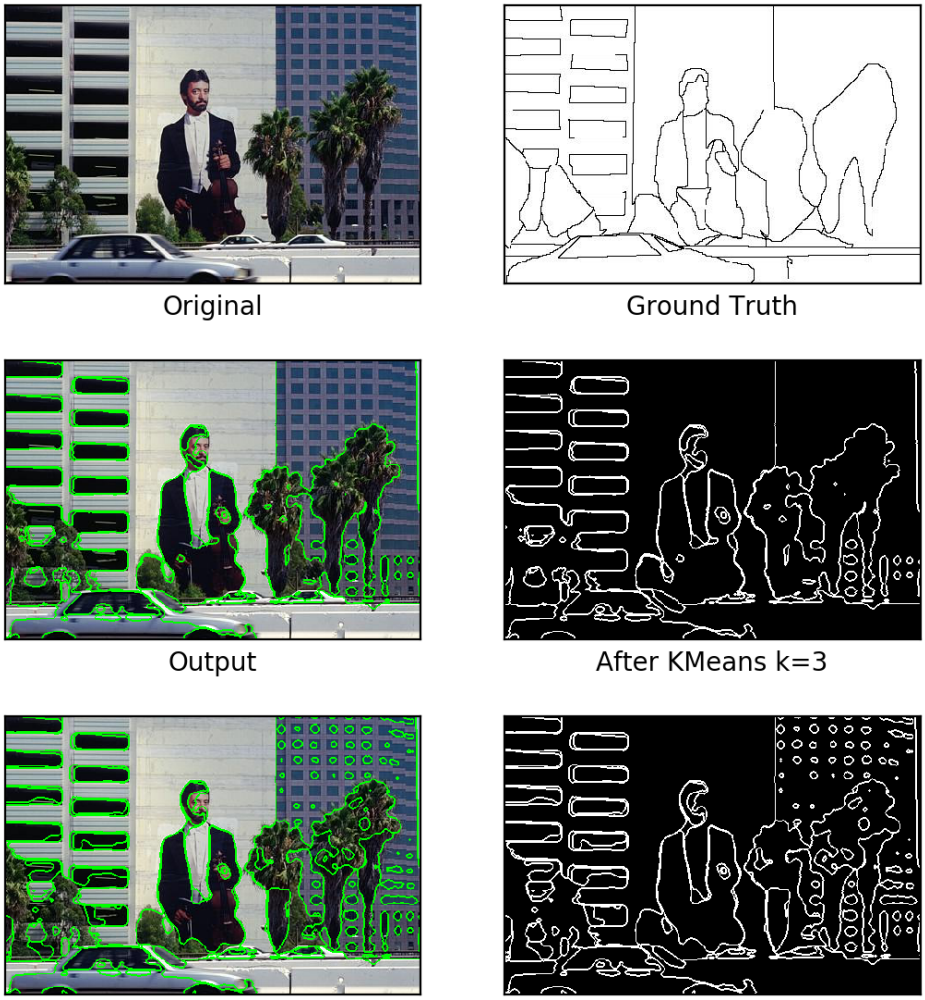
Original
Ground Truth
Output
After KMeans k=3
Output
After Mean Shift

Figure 3

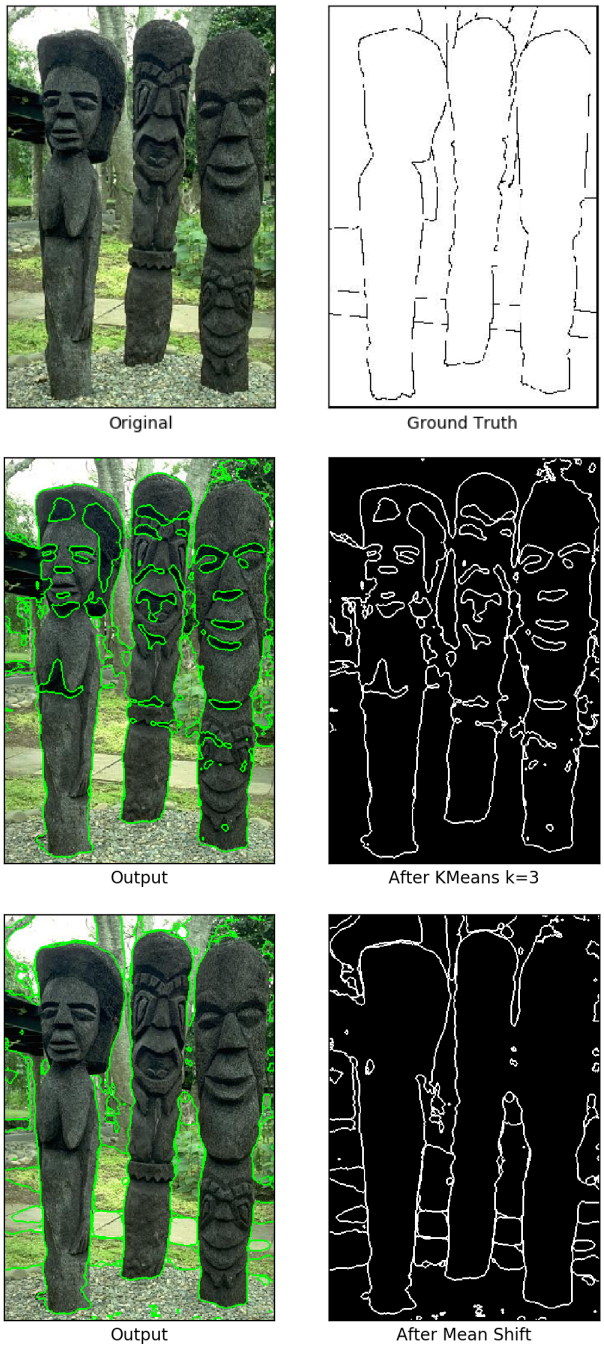
Original
Ground Truth
Output
After KMeans k=3
Output
After Mean Shift

Figure 4

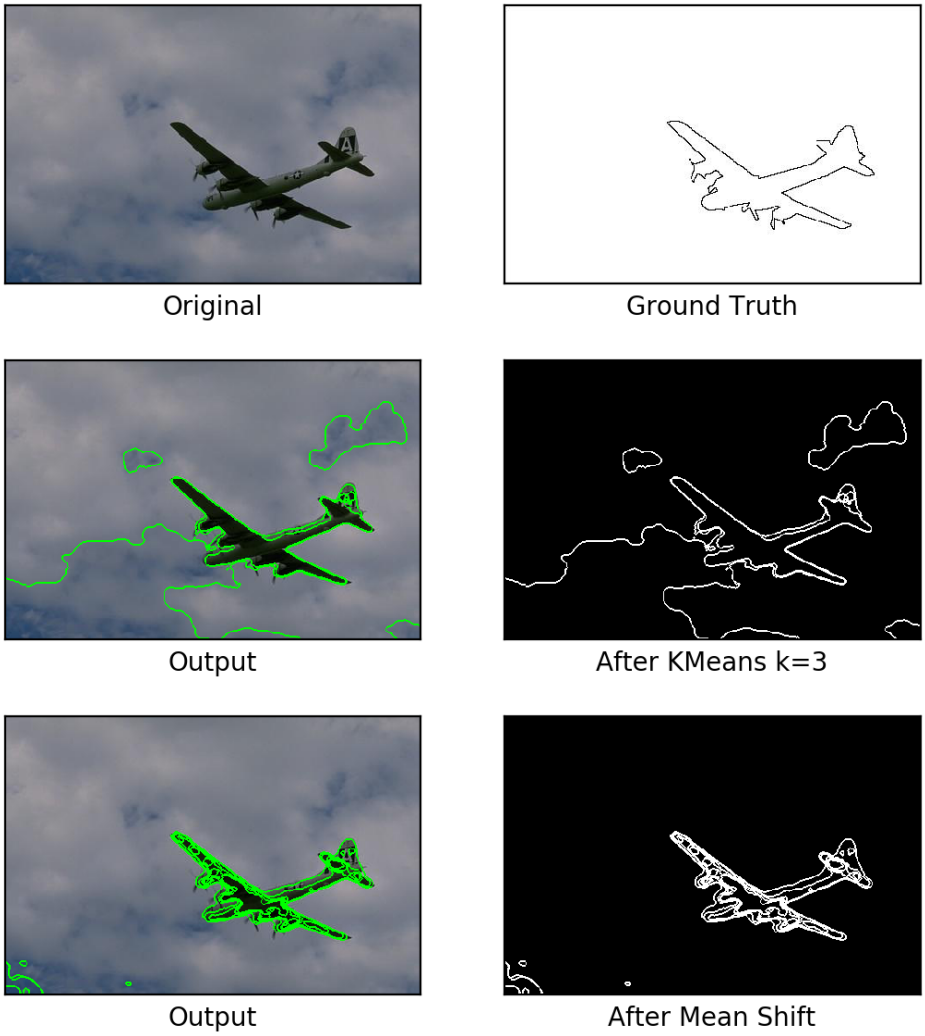
Original
Ground Truth
Output
After KMeans k=3
Output
After Mean Shift

Figure 5

A.3 Inference

- K-Means with k=2(or low value) consolidates images into the major colors. This helps is separating the foreground from the background.
- Mean Shift offers better segmentation in certain cases.
- Both methods fail when there is an excess of detail or when the color of the foreground and background are not very distinguishable.

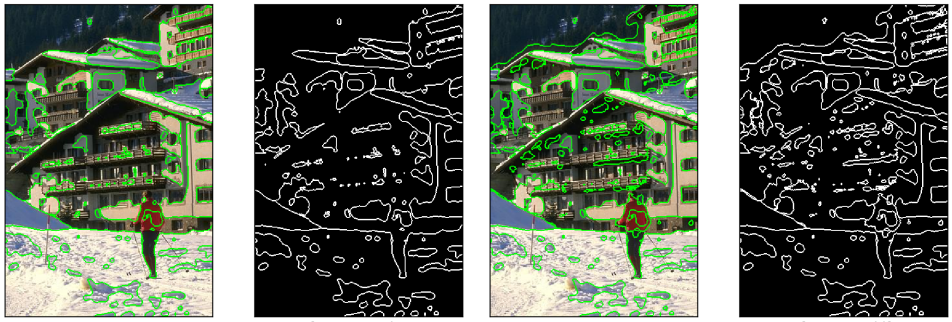

Figure 6: Where segmentation fails

B Panorama

Play with Panorama

Answers

B.1 Observations

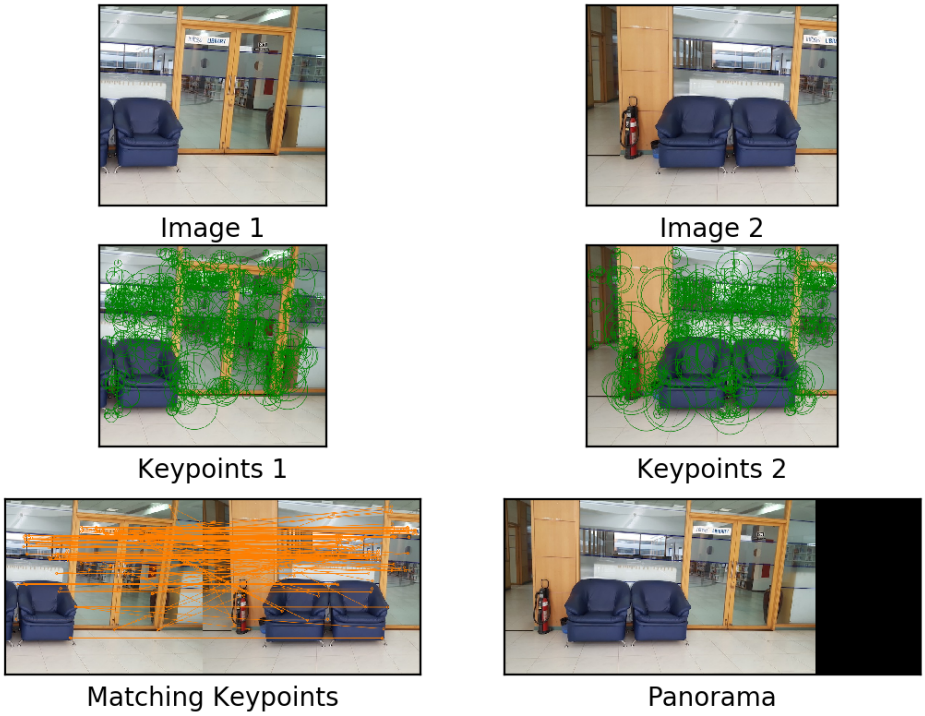

Figure 7: Panorama One

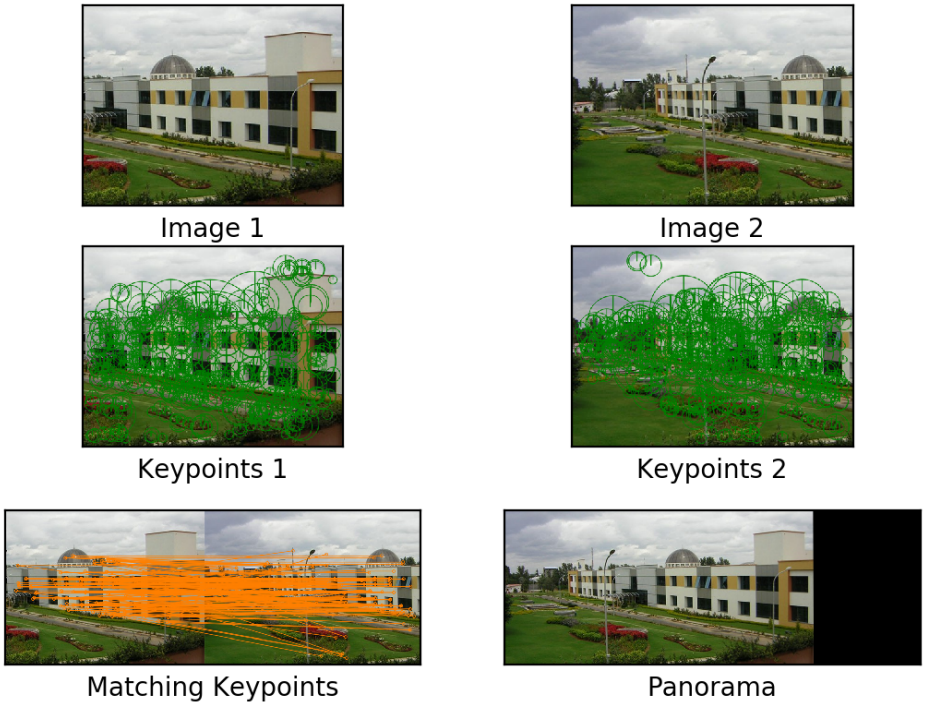

Figure 8: Panorama Two

B.2 Inference

- The images are stitched based on the matching keypoints found in left and right image.
- The images taken for stitching should be in the right order in which it is taken, if the order of stitching changes then the panorama creation fails.
- Wrapping of images might create black area due to the overlapping/stitching, which can be removed by cropping them.

B.3 SURF vs SIFT

`Explain how SURF is different from SIFT ? (10 sentences)`

Scale Invariant Feature Transform(SIFT)	**Speeded-Up Robust Features(SURF)**
SIFT method uses difference of Gaussian to build image pyramid.	SURF uses integer approximation to the determinant of Hessian blob detector.
SIFT we use different scales of image	SURF uses different scales of Gaussian masks while scale of image is always unaltered. Saves a lot of time.
SIFT uses an orientation histogram, and finds the largest orientation value and also those values that are over 80% of the largest which is used as the main orientation of the feature descriptor.	SURF uses the sum of the Haar wavelet response around the point of interest.
SIFT uses 128-D vectors for descriptors.	SURF extracts sum of absolute value of responses in x and y direction. The sum of the absolute values of the responses is also extracted. Thus each region has a 4-D descriptor vector. Concatenating everything, SURF has a 64-D vector which makes it faster.
SIFT uses approximated Laplacian of Gaussian with Difference of Gaussian for finding scale space.	SURF uses approximates Laplacian of Gaussian with Box filter. One Advantages is the convolution with box filter can easily be calculated with integral images which can be done parallely for different scales.
It is an iterative approach, which makes it slow.	Due to the use of box filters and integral images, we do not have to iteratively apply the same filter to the output of the previously filtered layer but instead can apply filter of any size at exactly the same speed directly on the original image and even in parallel. Therefore the scale space is analyzed by upscaling the filter size rather than iteratively reducing the filter size
It builds an image pyramid, filtering each layer with Gaussians of increasing sigma values and taking their difference	SURF creates a stack without 2:1 downsampling of higher values in the pyramid resulting in an image 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.

B.4 Principles of FLANN matching

`Briefly explain the main principles of FLANN matching (5 sentences)`

FLANN is a library of algorithms for performing approximate nearest neighbour in large dataset and for higher dimensional space.

Features of FLANN matching:

- FLANN uses two algorithms, randomized kd-tree and hierarchical k-means tree for computation of ANN (Approximate Nearest Neighbour).
- kd-trees are used for nearest neighbour computation. Randomized kd-tree algorithm is faster than conventional kd-tree algorithm (which is efficient for low dimensions, but slow in higher dimensions), as it uses multiple randomized kd-trees. The split dimension is chosen randomly from the first d-dimensions on which the data exhibits highest variance.
- Hierarchical k-means works by splitting the data points at each level in K distinct regions using k-means clustering. The same method is applied in turn to each region until number of points in a region is less than K.
- Best choice for ANN algorithm depends on features of the data such as structure and dimensionality of the data and the required search percision.
- FLANN considers the problem of choosing an optimal ANN algorithm as an optimization problem and runs it on a subset of the given dataset with weights for different parameters(tree build time, search time, memory overhead) based on which the cost is calculated.
- The use of automated algorithm configuration makes it easy to incorporate any new algorithms that are found in the future to have superior performance for particular datasets.

C Bikes vs Horses Classification

Implement bikes v/s horses classification.

Answers

Approach

- Application/System (Object Identification)
  - MP Module (Horse vs Bike Identification)
    - ML task (Binary Classification)
      - Features (Feature histogram of visual vocabulary terms created by clustering SURF/SIFT detectors)
      - Models (KNearest Neighbour or Logistic regression)
      - Optimization algorithm (eg: Least squares or gradient descent)

C.1 Procedure

Given data consists of bikes and horses. Part of the data was randomly kept aside for testing and rest was used for training.

Procedure 1

Using SIFT feature as Bag Of features + K-Means Clustering + FLANN Matching and KNN for classification

- Training data is labelled as Horse or Bike.
- The training images are read and the corresponding feature descriptors are extracted using SIFT feature detectors.
- The extracted descriptors are then added to Bag of Features, which is then clustered to extract relevant features using KMeans clustering with k=15 clusters.
- The dictionary returned from Bag of Features Cluster is then used as vocabulary. FLANN matcher which matches the descriptors of the training images with the nearest cluster centers to create feature histogram for each training image.
- The above list of labelled feature histogram is used to train the KNN classifier.
- Similarly the feature histogram is created for the test image which is used to predict the unknown label.

Procedure 2

Create a Bag Of Words class from scratch using SURF descriptors + K-Means Clustering + Logistic Regression.

- Detect keypoints and extract descriptors using SURF from each training image.
- Cluster the list of descriptors using K-Means to obtain the visual vocabulary.
- Create a feature histogram for each image using the visual vocabulary.
- Create a labelled dataset of feature histograms.
- Train a Logistic Regression model using the labelled dataset.
- For each test image, extract descriptors using SURF.
- For each descriptor find the nearest cluster and create the feature histogram.
- Feed the feature histogram into the Logistic Regression model to get the predicted label.

C.2 Observation

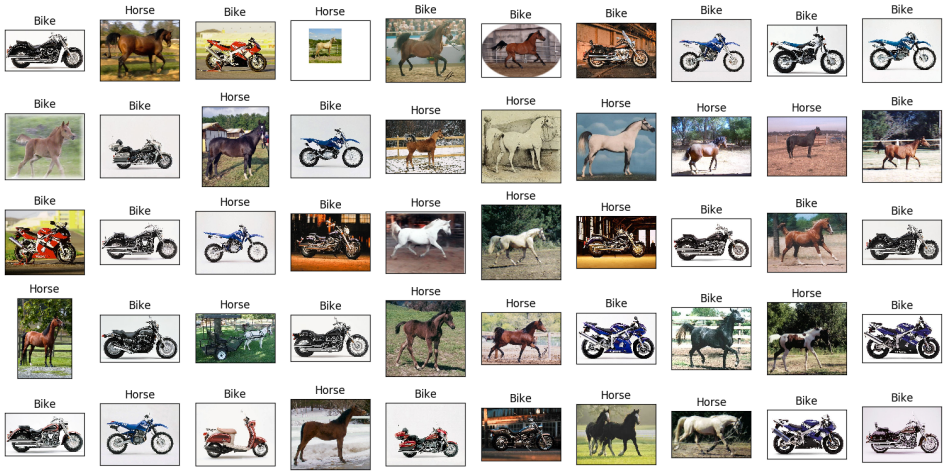

Figure 9: Classification Bikes vs Horses

C.3 General Inferences

- A peak in accuracy was found when the number of clusters was between 8 and 16. There was a slight decline and later a sharp decline as the number of clusters went over 100.
- The given dataset for bikes are skewed as it contains only the profile views of the bikes.

C.3.1 Inference 1

Training Image : 129
Number of clusters: 15
Testing Images : 50
Accuracy : 86% Accuracy

C.3.2 Inference 2

Training Image : 109
Number of clusters: 15
Testing Images : 70
Accuracy : 94.2% Accuracy

[0]Please refer the attached output files for details.