Solution 1

K-Means Clustering

k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Here, we use k-means to segment the input image. Randomly k cluster centres are chosen at the beginning and each point in image is allotted to cluster centre nearest to it. The mean of the cluster is then updated. The iteration termination criteria used by us is to stop if specified accuracy(i.e. 1 used by us) or the maximum number of iterations (i.e. 10 used by us).

We cluster into 2 clusters as we want to segment and find the foreground object from the image. This k-means can be performed on colour, intensity, or a combination of both. We run K-means algorithm to convergence, for any particular value of K(here 2), by re-drawing the image replacing each pixel vector with the {R, G,B} intensity triplet given by the centre μ_k to which that pixel has been assigned. Combining with the intensity in the approach followed by us does not give any additional better results.

We run the canny edge detector on this segmented image to compare with the ground truth. On doing so, we observe that there are lot of details being captured along with the segment. To reduce this excessive details available, we perform median blurring on the image before performing k-means clustering. This soothes out the details and thus a better region is captured by performing the canny edge detection.

Median shift algorithm

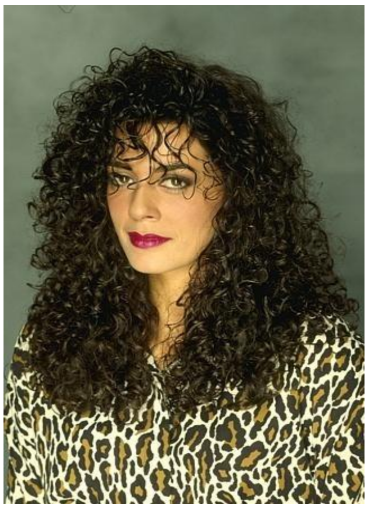
The mean-shift algorithm builds on the principle that there are a few modes in the image and rest of the pixels come from this mode. The features chosen in this case is the colour. Individual pixels are initialized and for each window, mean shift is performed till convergence. Near peeks are merged into one peek. This segmentation gives similar results to that of k-means.

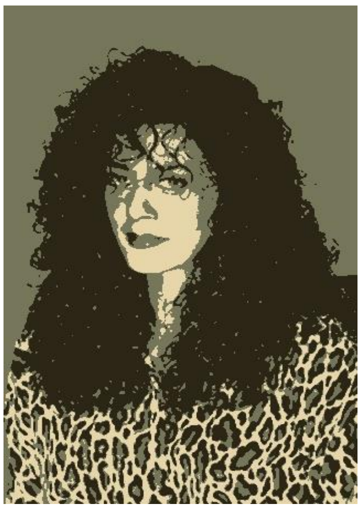
The criteria for iteration completion is similar to k-means, i.e. either till achieved accuracy or the maximum iterations are performed. The canny edge detection and median blurring is also performed in this case for comparison with the ground truth.

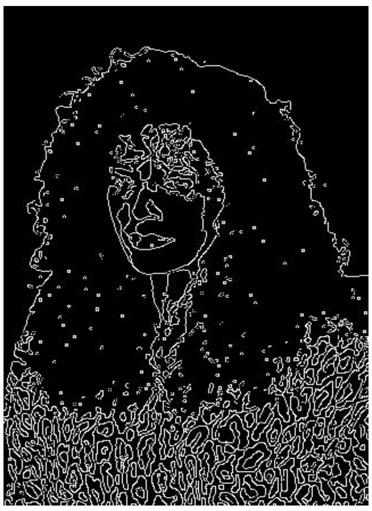
Observations

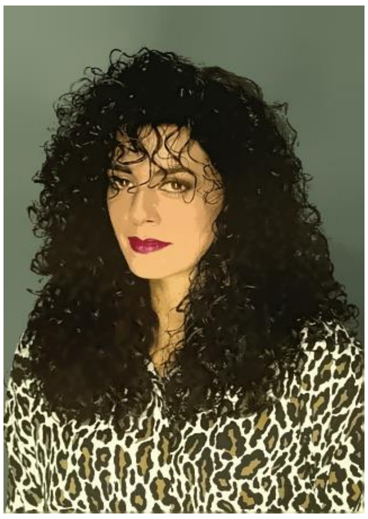
- We observe that sudden change in gradient is not approximated well.
- Thee texture affects the segmentation as there is difference in colour and intensity at the place of texture(e.g. hair or tree bark) and gets segmented into different segments. The blurring reduces a little of this issue but does not eliminate it.

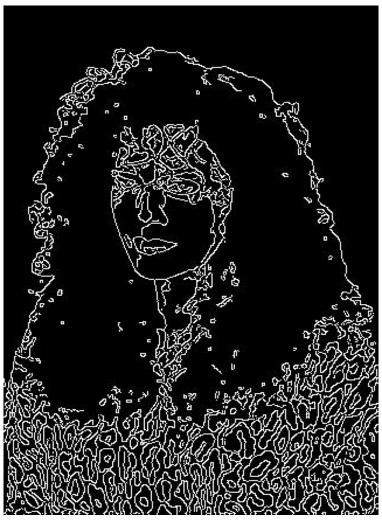
Output (Image 1):


Original Image


Kmeans(K =2)


Kmeans Boundary Image


Meanshift


MeanShiftBoundary

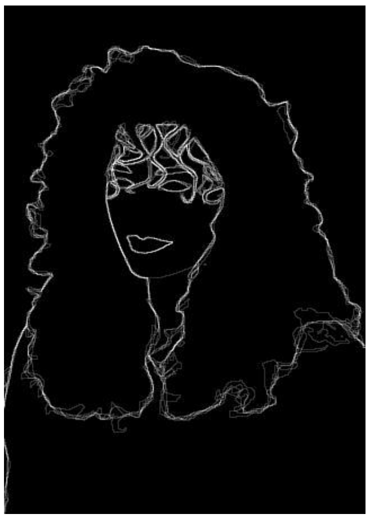
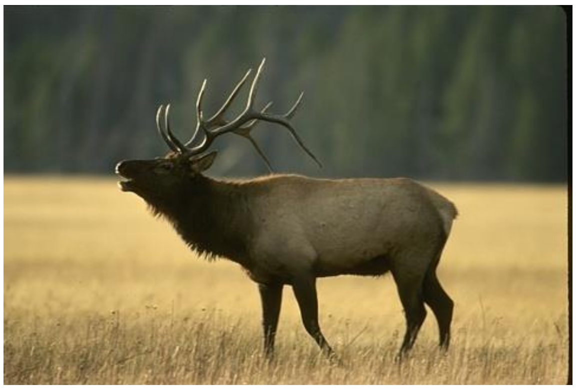
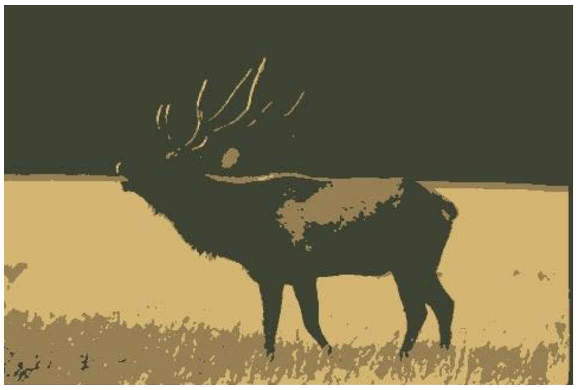
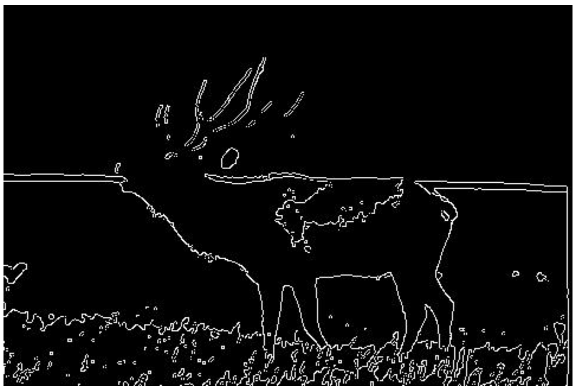
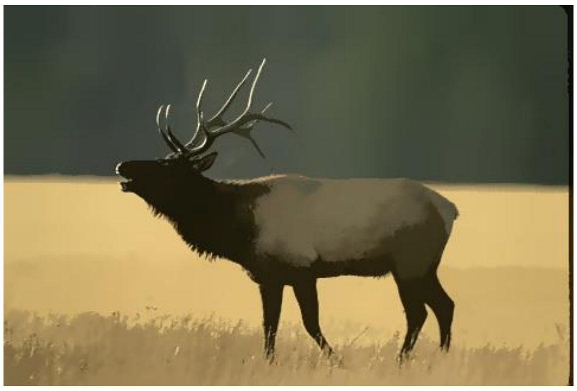

Ground Truth Image

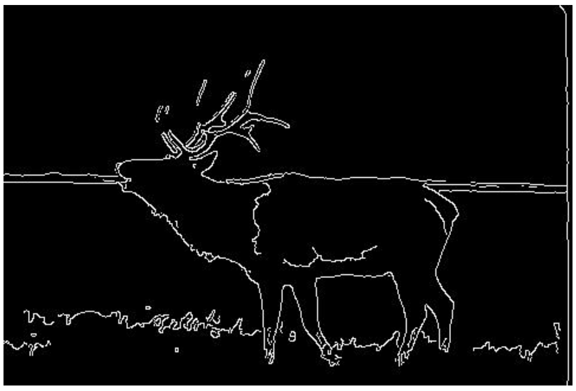
Image 2:


Original Image


Segmented Image(K =2)


Kmeans Boundary Image


Meanshift


Meanshift Boundary Image

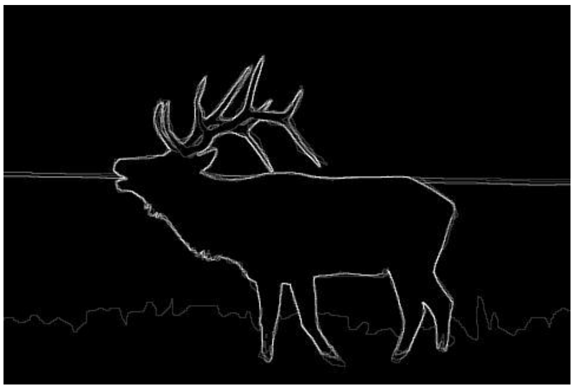
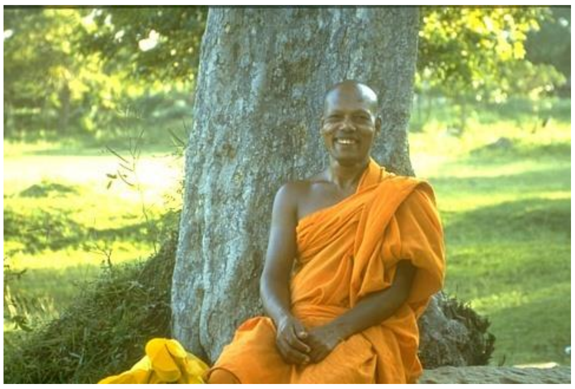
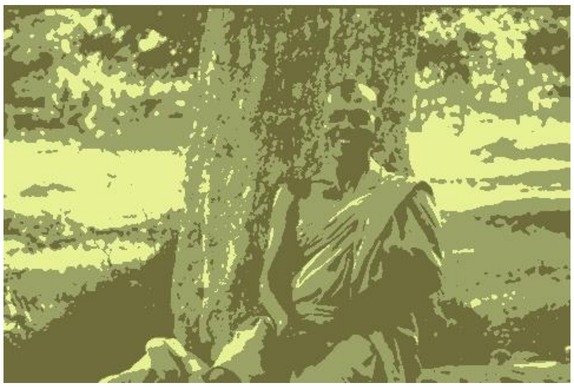
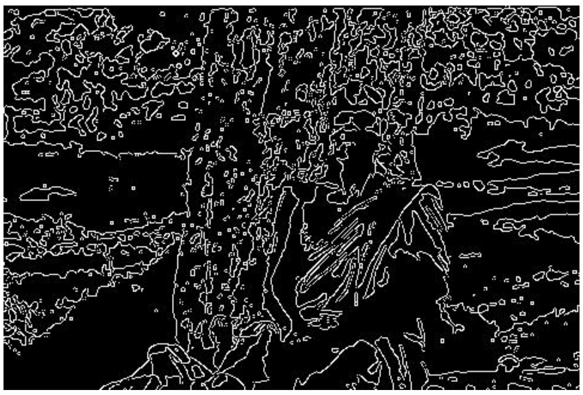
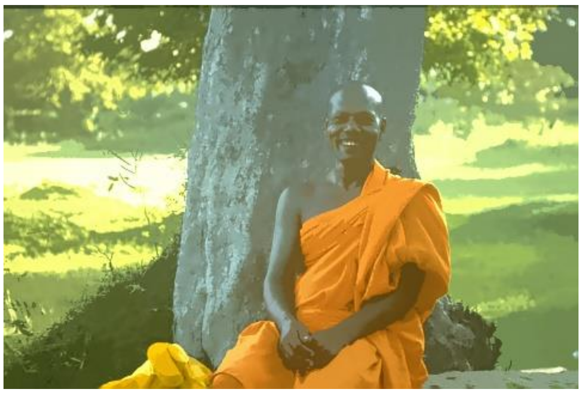

Ground Truth

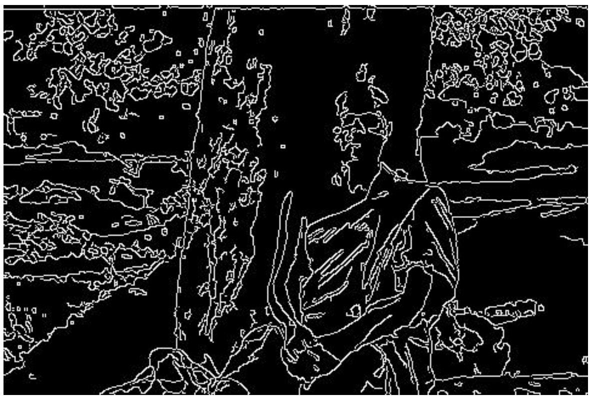
Image 3:


Original Image


Segmented Image(K =2)


Boundary Image


Meanshift


Meanshift Boundary Image

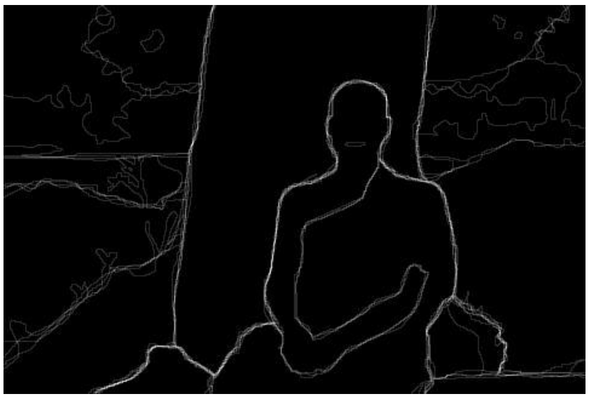
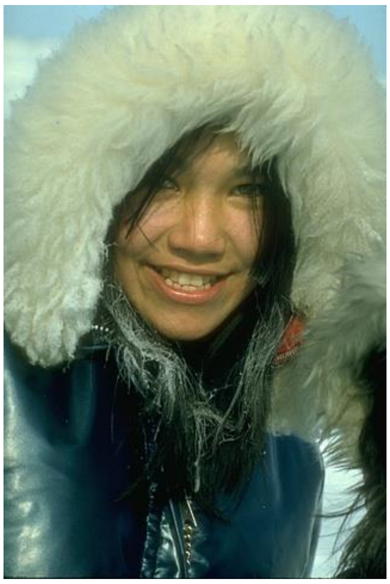
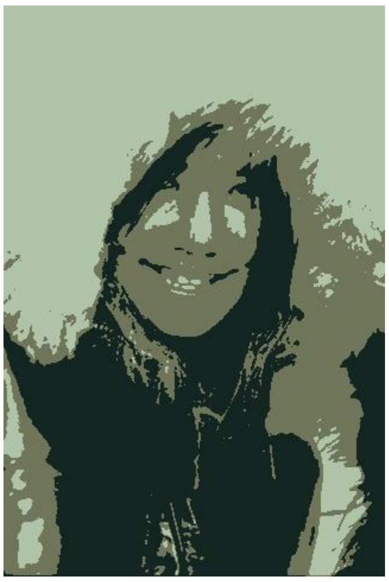
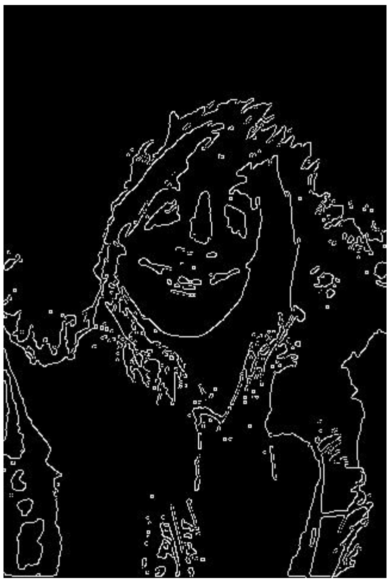
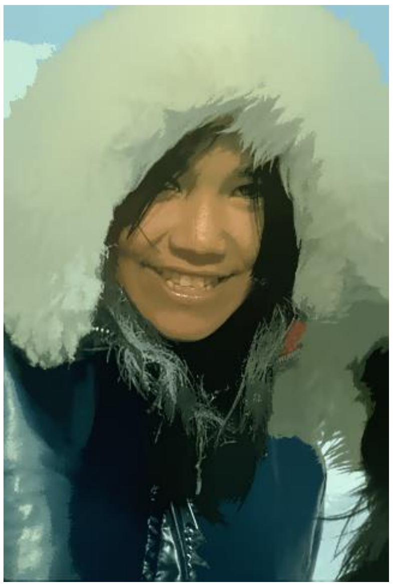

Ground Truth

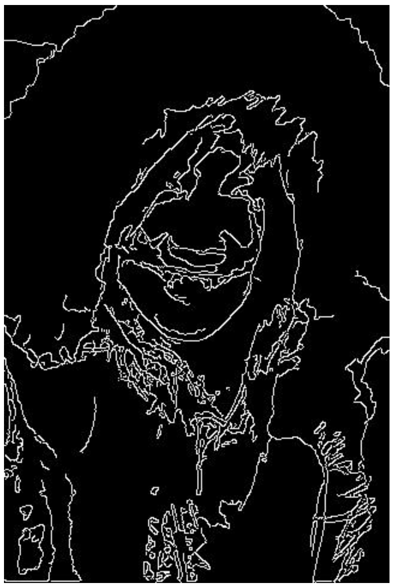
Image 4:


Original Image


Kmeans(K =2)


Kmeans Boundary Image


Meanshift


MeanShiftBoundary

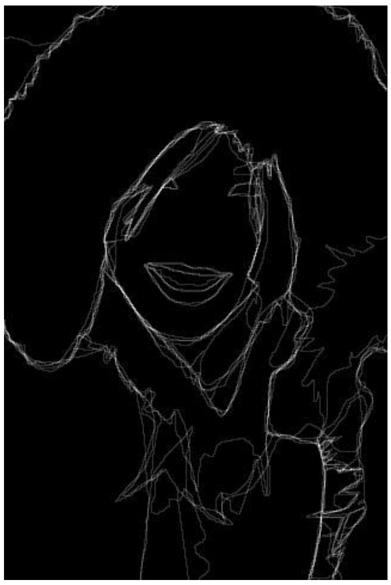
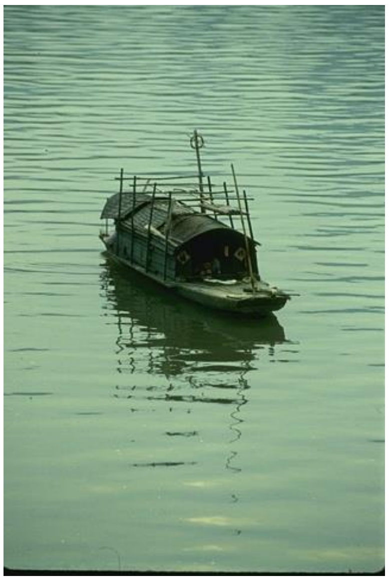
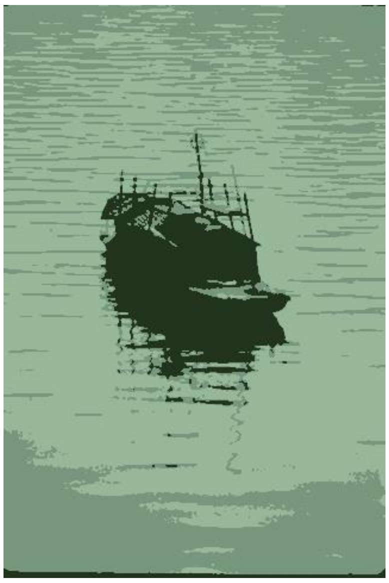
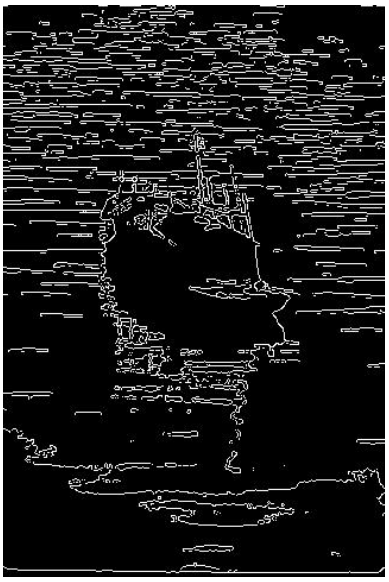
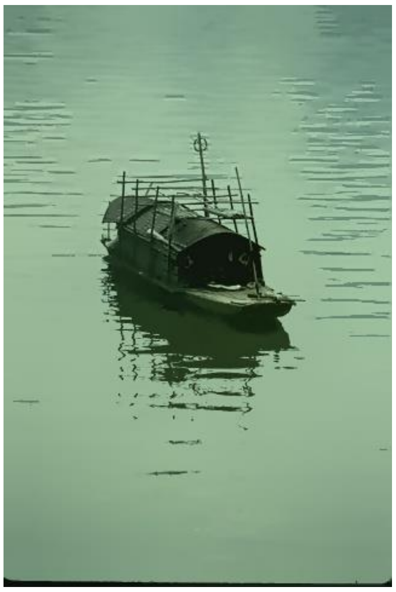

Ground Truth Image

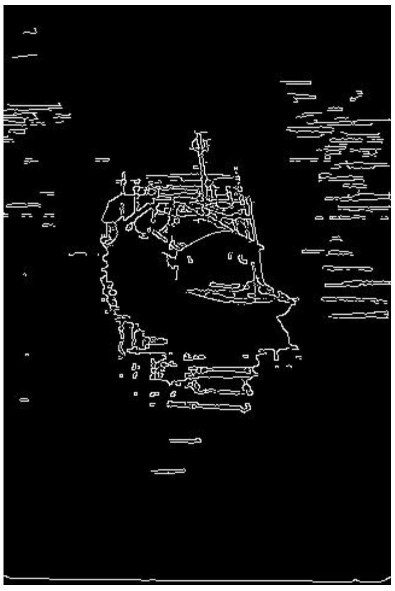
Image 5:

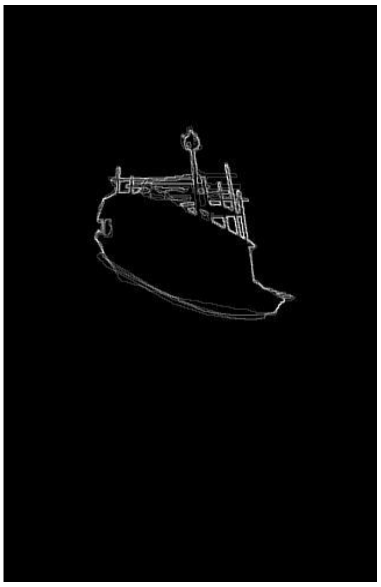

Original Image


Kmeans(K =2)


Kmeans Boundary Image


Meanshift


MeanShiftBoundary


Ground Truth Image

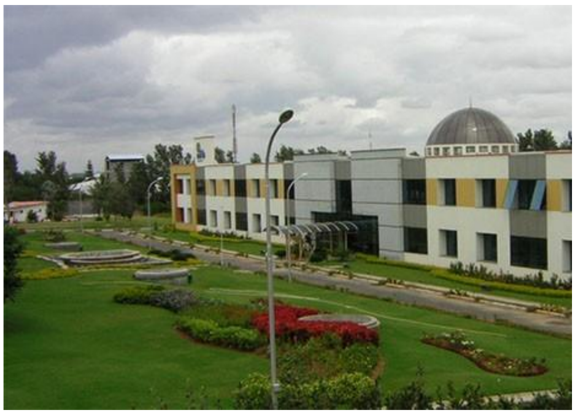
Solution 2

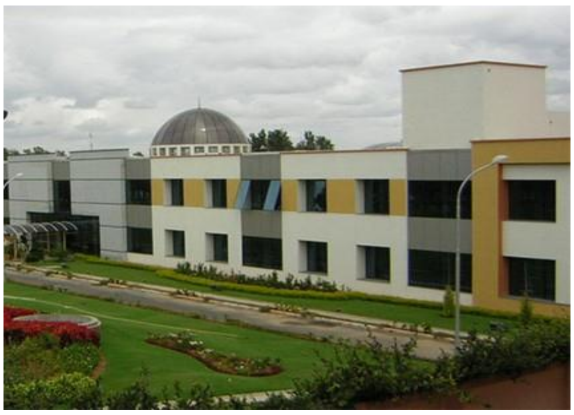
Panorama Stitching

We have edited the panorama stitching code provided according to python 3.1 where xfeatures is used rather than nonfree for the SIFT/SURF. The stitching output is attached below.

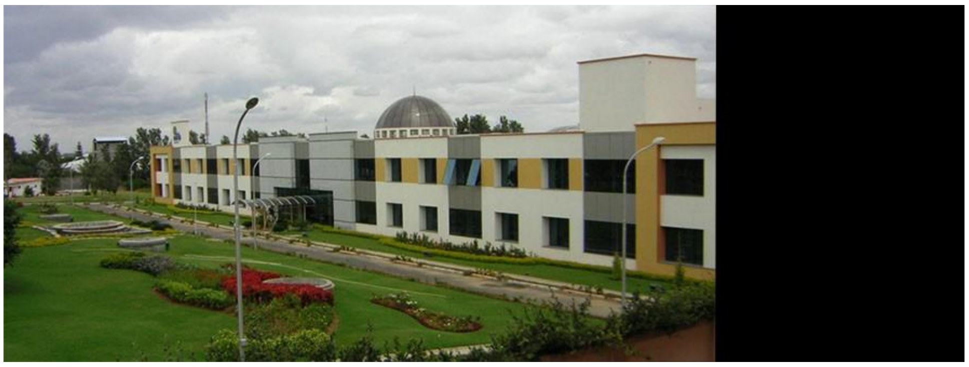
Results:

Input Image 1:

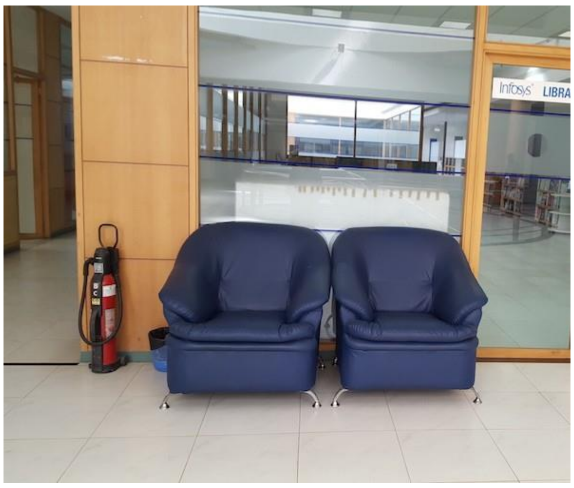


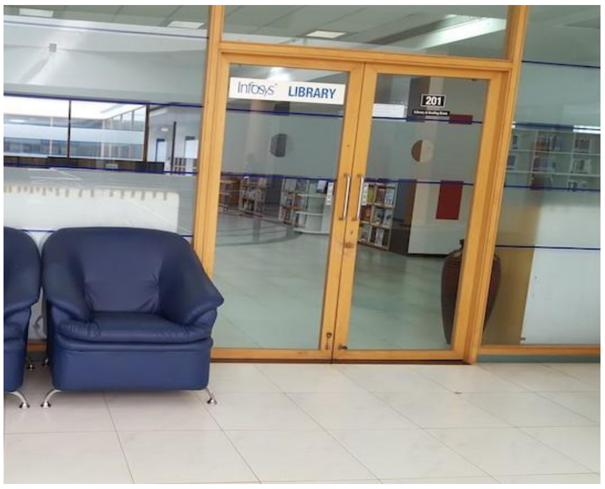


Output:

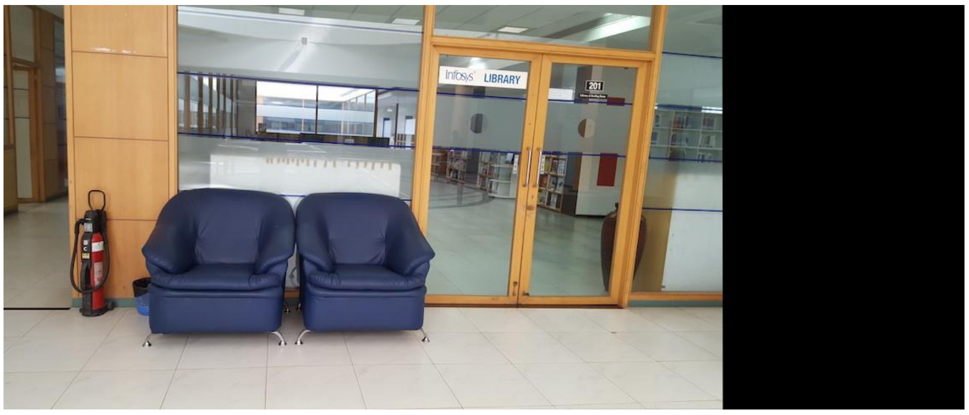


Second sample image:





Output:



SURF vs SIFT

- SIFT builds an 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.
- SIFT uses Difference of Gaussian to build image pyramid, but SURF uses integer approximation of determinant of Hessian blob detector. 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.
- In SIFT pyramid, we use different scale of image, while on the other hand, SURF uses different scales of Gaussian masks.
- ‘Fast-Hessian’ detector that used in SURF is more than 3 times faster that DOG (which was used in SIFT). Thus, SURF is faster than SIFT.
- To detect orientations, SIFT uses orientation histogram, while SURF uses sum of horizontal and vertical wavelet response around point of interest.
- SURF looks fast and good in most situations, but when the rotation is large, it also needs to improve this performance. SIFT shows its stability in all the experiments except for time, because it detects so many key points and finds so many matches.

Principles of FLANN matching

- For high-dimensional spaces, there are often no known algorithms for nearest neighbour search that are more efficient than simple linear search. An approximation of this has to be faster in order of magnitude, but obtain a near-optimal accuracy.
- The approximation algorithm should not be dependent on properties of the datasets, such as dimensionality, correlations, clustering characteristics, and size. FLANN(Fast Library for Approximate Nearest Neighbour) matching does automatic algorithm selection and configuration, which allows the best algorithm and parameter settings to be determined automatically for any given dataset.
- It performs a quick and efficient matching by using the clustering and search in multi-Dimensional Space modules.
- This matcher trains a model on a train descriptor collection and calls its nearest search methods to find the best matches.
- A cross-validation approach to identify the best algorithm and to search for the optimal parameter values to minimize the predicted search cost of future queries.

Solution 3

To perform classification, we consider the bag-of-words model. The bag-of-words model aims to identify the visual vocabulary based on the interest points in the image. For each interest point detected, we need to understand which word from vocabulary it represents. A histogram of such words contained in the image helps to determine the class of the object.

A detailed explanation of the approach used is:

- Find the SIFT descriptors for each image. These form the interest points.
- Now, to generate the vocabulary, we need to perform k-means over this set of interest points with k equal to number of words in the vocabulary. In our case k = 100.
- The image needs to be written in terms of the words from the vocabulary to be able to identify the histogram of visual words in the image. So for every image's interest point, we express it in terms of the k-means cluster centre to depict the word from vocabulary using the vector quantization.
- For each of the image obtained with histogram of visual words, we obtain the label for training.
- StandardScaler() is used to standardize features by removing the mean and scaling to unit variance. This is done as the system might behave badly if the individual features do not more or less look like standard normally distributed data
- The test images are also converted into histogram of visual words using same approach. The image thus obtained, is classified into appropriate label using the K-Nearest Neighbour approach.

We observe the following :

We took 65 images of bikes and 85 images of horses. Our test set had 35 images.

- This classification system has 91.43% accuracy for above mentioned dataset.
- Image with occlusion for the horse is also classified correctly.
- Images with multiple of horses or bikes are also correctly classified.
- The classification system misclassifies multiple bikes parked at an angle. This may have occurred as some interest points detected in that alignment have not been trained into the visual vocabulary, thus a proper histogram of visual words may not have been constructed leading to the misclassification.

Some outputs:

Incorrect classifications:

:

Some Correct Classifications: