VR Assignment – 2

By – Ajay Chhajed

(IMT2019006)

2.a)

Play with Panorama

Explanation

- Image stitching: It is a process in which we combine the images using key points.
- So, first we need to find the interest points. We will gather those points using SIFT, BRISK, ORB or SURF.
- After this, using KNN algorithm we will match these key points.
- Now, find Homography matrix using cv2.findHomography(src, dst, cv2.RANSAC, 5.0) and warp using cv2.warpPerspective(...).

SURF vs SIFT

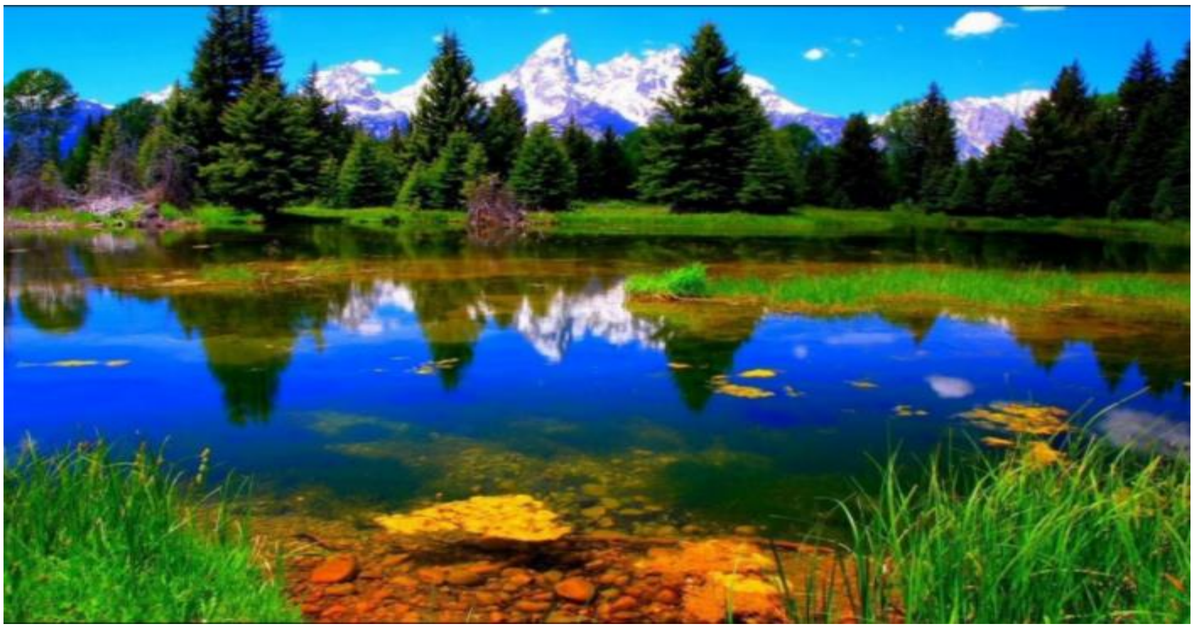
- SIFT (Scale-Invarient Feature Transform) and SURF (Speeded up Robust Features) both are algorithms used to extract features from the images.
- SURF is better than SIFT in terms of efficiency and speed.
- SURF reduces the computational complexity.

- SURF detects image key-points and generates descriptors using Hessian matrix while SIFT extract local features in images.
- SURF can give decent image result even when images are blurry while SIFT does not.
- SIFT can give decent image result even when images have different scales while SURF does not.
- SURF is 3 times faster than SIFT as SURF use squares for the approximation instead of gaussian average (of image).
- SURF is also better than SIFT in terms of rotation invariant and warp transform.
- SIFT matches its features from the image using Euclidian-distance based nearest neighbour approach.
- In SURF, convolution with box filter can easily be calculated using integral images or images formed after applying filters.

Left image

Right image

Output image



FLANN matching and RANSAC:

- RANSAC (Random Sample Consensus): It is an iterative algorithm for predicting parameters using observed data.

- RANSAC uses repeated random sub-sampling. The concept behind this is to use the subset of inliers for predicting parameters of the model.
- FLANN (Fast Library for Approximate Nearest Neighbours): It is a set of algorithms/techniques which are efficient enough to search the nearest neighbours in large dataset.
- FLANN uses data structure, <KD-Tree> for searching a nearest approximate neighbour and it works faster than brute force matching.
- RANSAC uses minimal number of points to predict the model parameters whereas FLANN is advantageous enough only when number of descriptors are greater than 1K.

2.b)

<u>**Bike vs Horse Classification & CIFAR-10**</u>

Procedure

- BOVW (Bag of Visual Words): It is similar to BOW but in this case instead of words we have image pixels/features.
- We are using SIFT, SURF, BRISK, ORB to extract the features of the image. Now, these algorithms will give us the interest points and descriptors of the image using cv2 library, which we will use to construct vocabulary using k-means algorithm.
- Now features of the image will be represented in the form of histogram of frequency.
- These histograms will be used to train our model, in our case it is Logistic regression, as this model gives us the best result.

Approach

- Extract all the features and descriptors using algorithms SIFT, SURF or ORB.
- Apply K-mean on the descriptors to find vocabulary centres, for given images.
- Build frequency histograms of features using descriptors count in vocabulary.
- Train our model using these frequency histograms.
- Predict on the test data using model Logistic Regression.

Observations for Bike vs Horse

The best accuracy I got with k = 80 BOVW + SIFT is 0.94 using Logistic Regression.

Observations for CIFAR-10

Model	**Accuracy** (0 <= p <= 1)
BOVW + SIFT: k = 25	0.272
BOVW + SIFT: k = 50	0.280
BOVW + SIFT: k = 100	0.27

References

https://towardsdatascience.com/image-panorama-stitching-with-opencv-2402bde6b46c

https://www.cs.toronto.edu/~kriz/cifar.html

https://www.pyimagesearch.com/2016/01/11/opencv-panorama-stitching

https://towardsdatascience.com/bag-of-visual-words-in-a-nutshell-9ceea97ce0fb

https://medium.com/analytics-vidhya/bag-of-visual-words-bag-of-features-9a2f7aec7866

https://github.com/vatsal199/Bike-vs-Horse-classification/blob/master/CLS.ipynb -- for load data set