

Visual Recognition

Assignment - 2

Shubham Gupta

IMT2016118

IIITB

Assignment 2a

Play with Panorama:

In this we have stitched together a Panorama of an Institute, Outside Library of the Institute.

Goals

1. To get the panorama of the institute using two given images of institute.
2. To get the panorama of the outside library of the institute using two given images of outside library.

Procedure

To get the panoramic image we basically have to matching the common points of the given images and then we have to concatenate them to get the final image as a panoramic image. Because given images are going to get concatenate on common points between the images.

Approach

1. First read the the two given images of and institute and two given images of outside library.
2. "sift" function is used to find the key point and descriptor of image, like nose, eyes, lips etc in face image.
3. Matched the descriptor of two images and check the common part in two images using "cv2.BFMatcher" function.
4. After matching the descriptor of the two images, we have to sort the descriptors on the basis of distance and then the match points of image 1 and image 2 should be joined by a straight line.
5. Fill the common point of both images in array and then we find the matrix of transformation bw src_pts1(img1) and des_pts1(img2)
6. Apply perspective transformation to both images, to get the warped image.
7. Concatenate of 1st image(img1) and warped image to get the panoramic image.

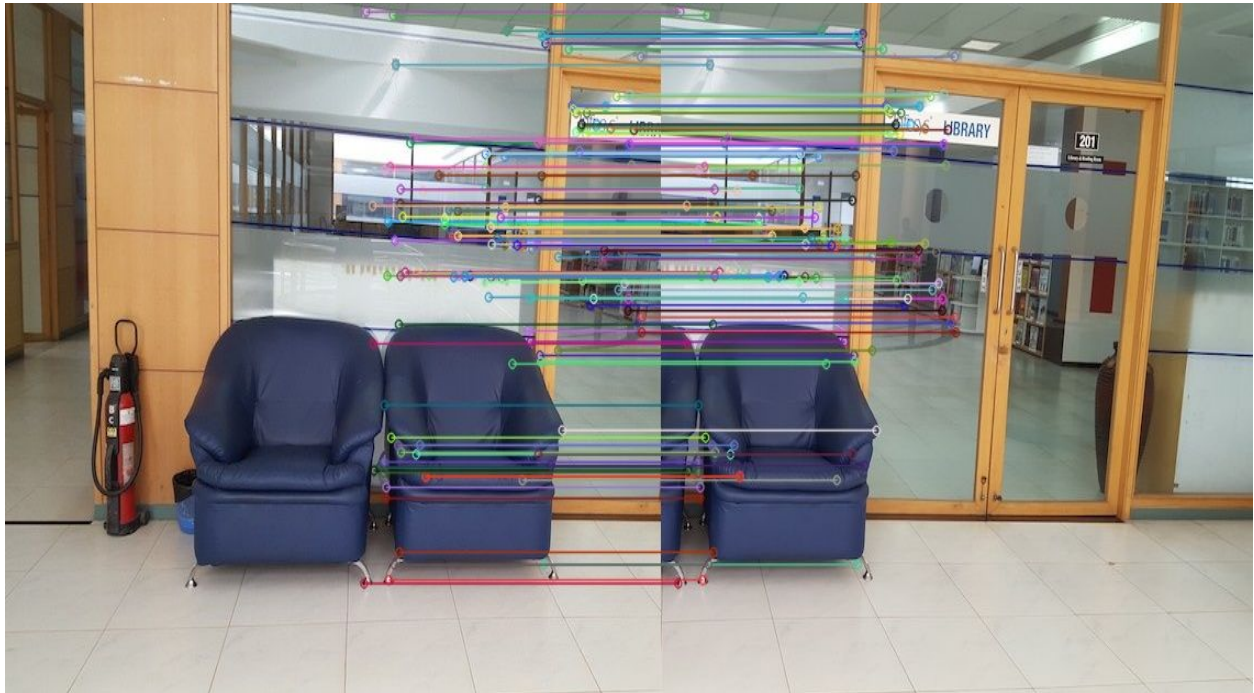
Observation



Key points image of Outside Library



Key point images of Institute



Outside Library image after matching points



Institute image after matching points

Panoramic Images



Outside Library Panoramic Image



Institute Panoramic Image

Difference between SURF and SIFT

1. Descriptors size in SURF is 64 bits while in SIFT it is 128 bits.
2. In general, robustness of SURF is same as that of SIFT, but depending on the datasets we may sometimes get SURF to be little less robust than SIFT.
3. SIFT detects more features than SURF, but SIFT has to compromise in speed.
4. In SURF, Laplacian of Gaussian (LoG) is convoluted with the integral image. While in SIFT, Difference of Gaussian (DoG) is convoluted with different size of images.
5. SURF performs better than SIFT in terms of matching feature pairs in rotation.
6. SURF provides better results than SIFT in matching feature pairs in different hues and different warps transformed.
7. SIFT performs better than SURF in terms of different scale, saturation and image value.
8. SURF is found to be 3 times faster than SIFT.
9. Orientation Assignments of SURF is different from SIFT.
10. SURF provides better results than SIFT in matching feature pairs in blur images and noisy images.

Principles of RANSAC matching

1. The RANSAC algorithm is a learning technique to estimate parameters of a model by random sampling of observed data.
2. RANSAC, attempts to exclude the outliers and find a linear model that only uses the inliers in its calculation.
3. RANSAC uses the voting scheme to find the optimal fitting result.
4. RANSAC selects a random subset of the original data. Call this subset the *hypothetical inliers*.
5. In RANSAC algorithm model is fitted to the set of hypothetical inliers.

Assignment 2b

Implement Bike vs Horse classification:

Goal

A data set of images of Horse and images of Bikes is given to us now we have to do the classification of Bike and Classification of Horse.

Procedure

We have to do the image classification using bag of visual words approach. For that first we have to do the Feature extraction of the given dataset and then we use different different classification model to the the maximum accuracy of classification.

We use the keypoints and descriptors to construct vocabularies and represent each image as a frequency histogram of features that are in the image.

From the frequency histogram, later, we can find another similar images or predict the category of the image.

Approach

- **Feature Extraction**
 - Features of images of Horse and Bikes are extracted using SIFT detectors.
 - To make the cluster from the image descriptors(50 in case of Horse and Bikes) we used K- means clustering algorithm and Finally, for each image, we make frequency histogram from the vocabularies and the frequency of the vocabularies in the image.
- **Building Model**
 - So basically the histogram that we obtained is used as a feature vector for every image and gets trained using various classification algorithms like SVM, K-Nearest Neighbour, Logistic Regression etc.

- **Result**

- I get the best result for Horse and Bikes classification when I used SVM(Support Vector Machine) and Logistic Regression with accuracy score of "0.9130434782608695".
- When I trained the model for the CIFAR-10 dataset K-Nearest Neighbour gave very less accuracy. But SVM turned out to give a very good accuracy score of "0.183478". Although which is also very less.

Observation

- Although we used same model for the CIFAR-10-python dataset the accuracy comes out to be very less mostly because the number of key points for each image in CIFAR-10 data set was very less , close to 5-6 key points per image in the dataset but in case of Bike and Horse dataset accuracy comes out is very good because no of key-points in those datasets is close to 1000-1800. This is the reason for not getting the good accuracy in CIFAR dataset.