Visual Bag of Words and Mosaic Creation

Skand Vishwanath Peri
2014CSB1034

Department of Computer Science & Engineering,
IIT Ropar

This assignment consists of 2 parts. The first part consists of the Visual Bag of words implementation using the k-means clustering and the second part consists of the mosaic creation using homography estimation and blending.

1 Visual Bag of Words

Visual Bag of Words consists of the following steps:

- Extracting Interest points
- Extracting features of patches around the interest points
- K-Means Clustering on the extracted features
- Creating Histogram for the training and testing set
- Test time prediction using histogram matching

1.1 Extracting Interest Points

I have used the Harris Interest (corner) points for extracting the interest points.

1.2 Extracting features of patches around the interest points

I have used HOG (Histogram of Gradients) features of the patches. Patches of 3×3 are cropped and hence the length of the feature vector (HOG feature) is 36. These features are extracted for all the training as well as the testing images.

1.3 K-Means Clustering on the extracted features

K-Means clustering is applied on these extracted features. Since we don't know what the value of k is so I find it empirically by using the **Elbow Method** which states that, to have an optimal clustering, choose that k whose ratio between-group variance to the total variance doesn't change for 2 consecutive clusters.
The complexity of K-Means clustering increases as the value of k increases. So I tried my experiments till $k = 250$ and decided to take $k = 200$ since not much change was observable in the neighbourhood. It might be the case that if we do the clustering for some other $k > 250$, we can get better accuracy but due to computational restrictions, I did not try for $k > 250$.

1.4 Creating Histogram for the training and testing set

This task includes creating histogram for each training and the testing set. Each image has some f features (interest points), so we take each one of these f features and compute the euclidean distance between each of the k clusters and then add the inverse of all the histogram values to get the final histogram of an image. This procedure is repeated for each of the training and the testing image. We take inverse since the distance is inversely proportional to the weight of that cluster.

1.5 Test time prediction using histogram matching

During the test time, we simply compare the histograms of each test image with all the training images and assign that label whose distance is minimum in the training set. This is how we classify an image during the test time.

2 Mosaic creation using homography

This section explains the method of creating a mosaic using homography. This techinique can be used in creating panorama images. It includes the following steps:

- Selecting Points

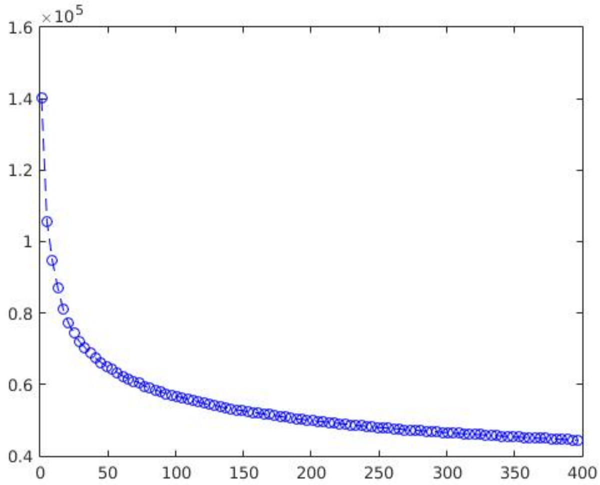

Figure 1: The elbow function on y-axis vs the number of clusters on x-axy

- Estimating Homography
- Warping images
- Blending the warped images

2.1 Selecting Points

We are trying to estimate a **projective transformation** and not affine transformation meaning that parallalism is not preserved in projective transformations. So we need to select at least 4 points for computing the homography matrix. For selecting the points I have used *getpts* function.

2.2 Estimating Homography

We have 4 points (8 coordinates) and need to estimate a 9 element homography matrix H such that:

$$\begin{pmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} = \begin{pmatrix} x'_1 \\ x'_2 \\ 1 \end{pmatrix} \tag{1}$$

where $\begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}$ is the point in image 1 and $\begin{pmatrix} x'_1 \\ x'_2 \\ 1 \end{pmatrix}$ is the corresponding point in image 2. These equations on futher solving can be boiled down to homogenous system of equation where:

$$PH = \begin{bmatrix} -x_1 & -y_1 & -1 & 0 & 0 & 0 & x_1x'_1 & y_1x'_1 & x'_1 \\ 0 & 0 & 0 & -x_1 & -y_1 & -1 & x_1y'_1 & y_1y'_1 & y'_1 \\ -x_2 & -y_2 & -1 & 0 & 0 & 0 & x_2x'_2 & y_2x'_2 & x'_2 \\ 0 & 0 & 0 & -x_2 & -y_2 & -1 & x_2y'_2 & y_2y'_2 & y'_2 \\ -x_3 & -y_3 & -1 & 0 & 0 & 0 & x_3x'_3 & y_3x'_3 & x'_3 \\ 0 & 0 & 0 & -x_3 & -y_3 & -1 & x_3y'_3 & y_3y'_3 & y'_3 \\ -x_4 & -y_4 & -1 & 0 & 0 & 0 & x_4x'_4 & y_4x'_4 & x'_4 \\ 0 & 0 & 0 & -x_4 & -y_4 & -1 & x_4y'_4 & y_4y'_4 & y'_4 \end{bmatrix} \begin{bmatrix} h1 \\ h2 \\ h3 \\ h4 \\ h5 \\ h6 \\ h7 \\ h8 \\ h9 \end{bmatrix} = 0 \tag{2}$$

The solution to this can be efficiently computed by the Singular Value Decomposition of the matrix P as $P = USV^\top$ and the last column vector of the matrix V is the homography matrix.

2.3 Warping images

After the homography is estimated, it is warped on the original image to get the transformed image. This has been done by using the inbuilt MATLAB command *imwarp*.

2.4 Blending of image

Now we have the 4 points in both the images and we stich the image 1 and the warped image 2. Then we blend the images with the coefficient α varying from 1 to 0 gradually. This is similar to the blending in the hybrid images[2]. Figure 2 shows an example of mosaic image.

3 Results & Discussion

For the first task i.e Visual Bag of words, the first challenge was to determine k. The plot of the sum of squared errors within each cluster vs the number of cluster is shown in Figure 1. The accuracy of the classification for $k = 200$ is 63%. The class wise accuracy, precision and recall is given as output of the MATLAB code. For the second task, the output is shown in Figure 2. The image is blended using the hybrid images concept. The image as been taken from [1].

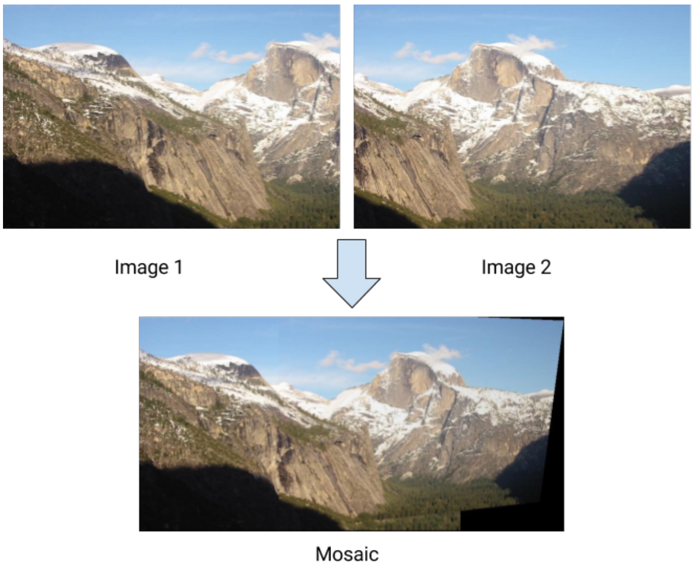

Figure 2: Image after the homography estimation, warping, stitching and blending.

The nearest neighbouring patch of every cluster is also stored for ease and visualization. Figure 3 shows the visualizations of those patches.

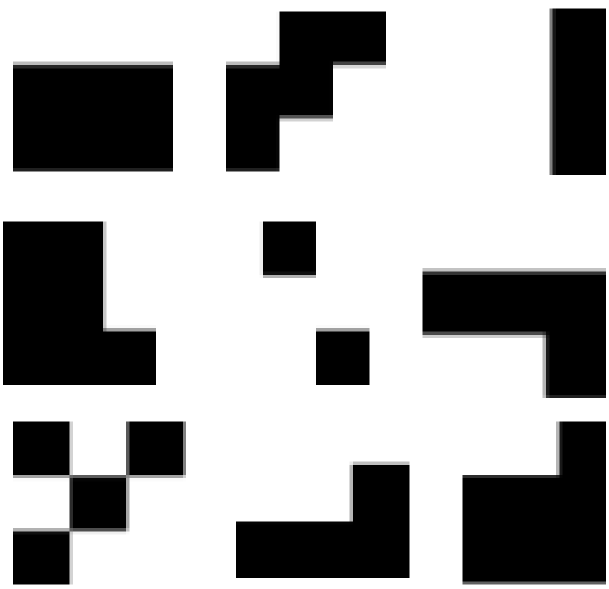

Figure 3: Visualizations of the nearest neighbour patches to the mean of the cluster.

4 References

[1] URL `http://www.cs.cornell.edu/courses/cs6670/2011sp/projects/p2/project2.html`.

[2] P.G. Schyns A. Oliva, A. Torralba. Hybrid images. *ACM Transactions on Graphics, ACM Siggraph*, pages 527–530, 2006.