Miniproject: Brief Report
(ECE 4039- Motion and Geometry Based Methods in Computer Vision)

Name(s): Shashank Shantam, Prachi Sinha, Tanishq Kandhari
Registration Numbers: 140907716, 140907604, 140907035

Title: Content Based Image Retrieval System

Objective: The objective of the project is to build an image search engine which searches in a locally stored database for images that are similar to the query image submitted. The system defines similarity based on the color distribution of the image which is quantified using 3D HSV Color Histograms.

Dataset used: We have used the *INRIA Holidays Dataset* for our dataset of images. This dataset consists of photos taken from various vacation trips from all over the world, including a very large variety of scene types (natural, man-made, water and fire effects, etc) and the images are in high resolution. The dataset contains 500 image groups, each of which represents a distinct scene or object. The first image of each group is the query image and the correct retrieval results are the other images of the group.

Dataset size: 1491 images in total: 500 queries and 991 corresponding relevant images

Number of queries: 500 (one per group)

Approach:

The entire approach is broadly divided into four major steps:

1.Defining the image descriptor - Our image descriptor is a 3D color histogram in the HSV color space. We use regions-based histograms rather than global-histogram because it allows us to simulate locality in a color distribution. The histogram is normalized to obtain scale invariance.

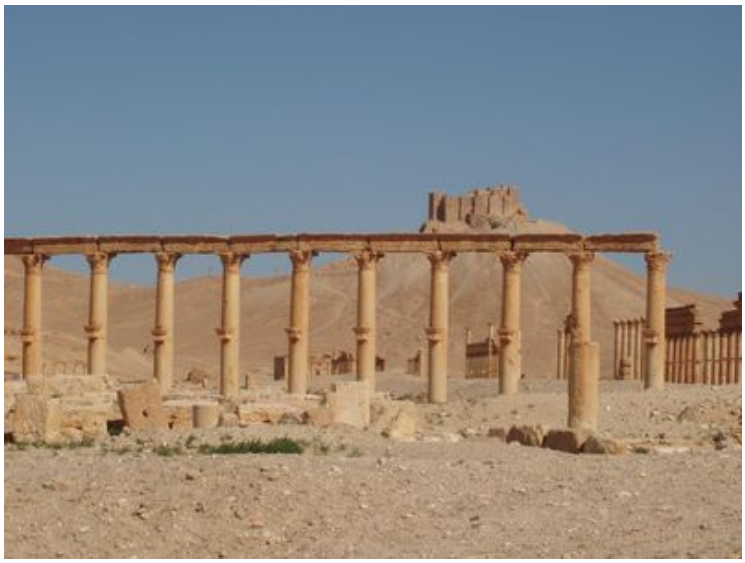
2.Indexing the dataset - Once we have our image descriptor defined, we extract features from each image in our dataset. This process of extracting features and storing them on persistent storage is called indexing. The extracted features are then stored in a CSV file

3.Defining the similarity metric - For our project, we use the Chi-Squared distance measure to compare the histogram of the query image with the dataset. This distance measure is used to quantify the similarity between two images: the smaller the distance, more alike the images will be.

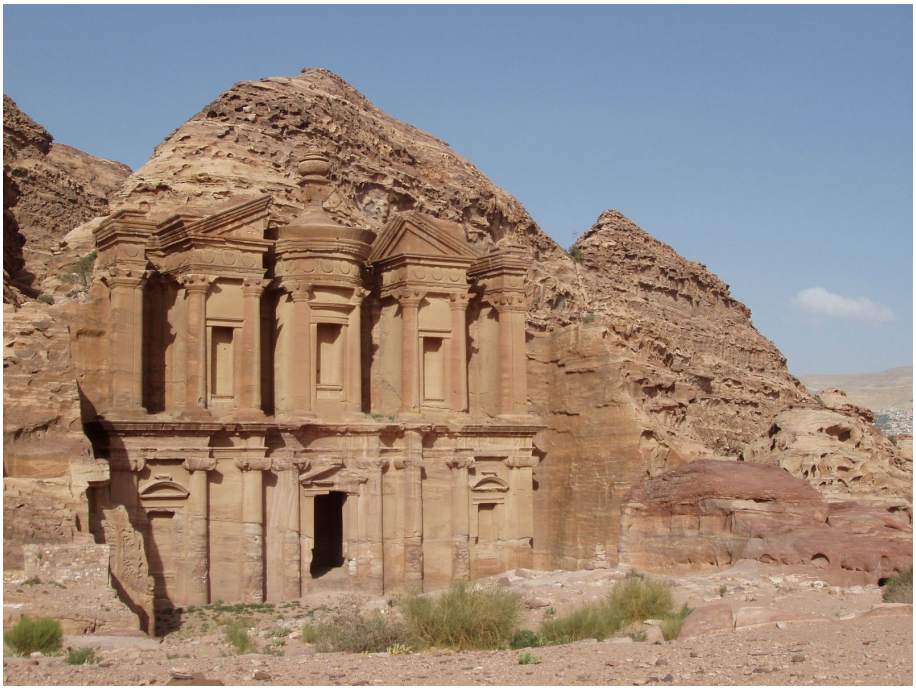
4.Performing the search - The final step is to perform the actual search. The user submits a query. The system then extract features of the query image, applies the chi squared similarity function to compare the query to the features already indexed. From there, we simply return 5 images with the smallest distance which corresponds to the most similar images.

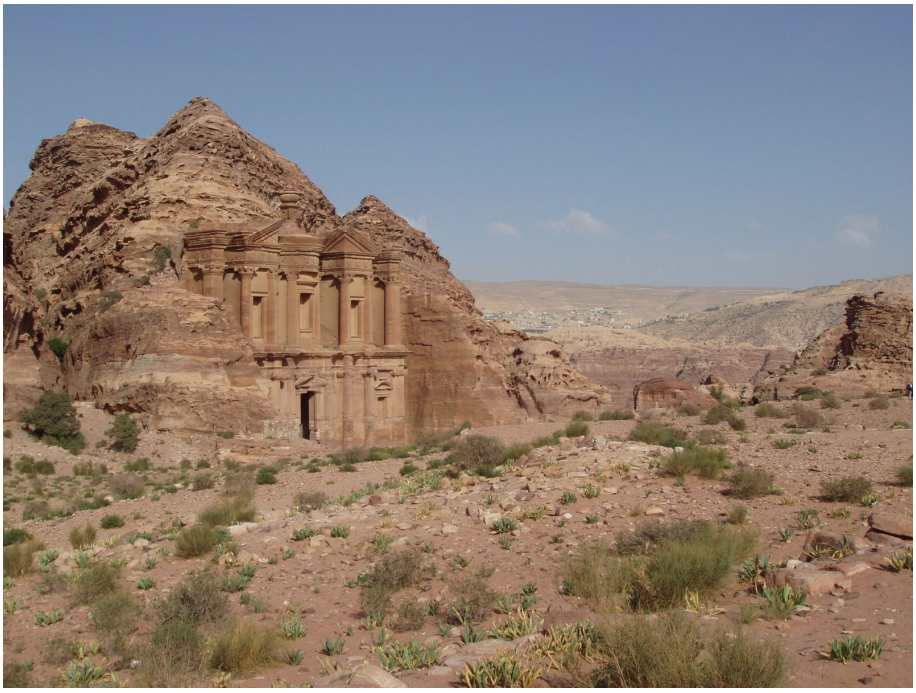
OUTPUT

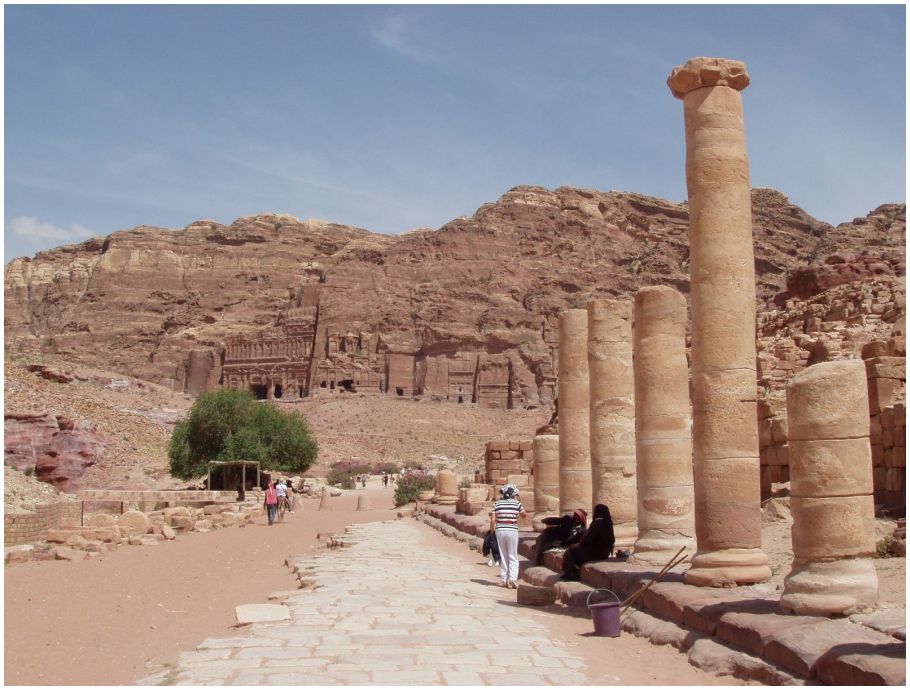
Query Image:

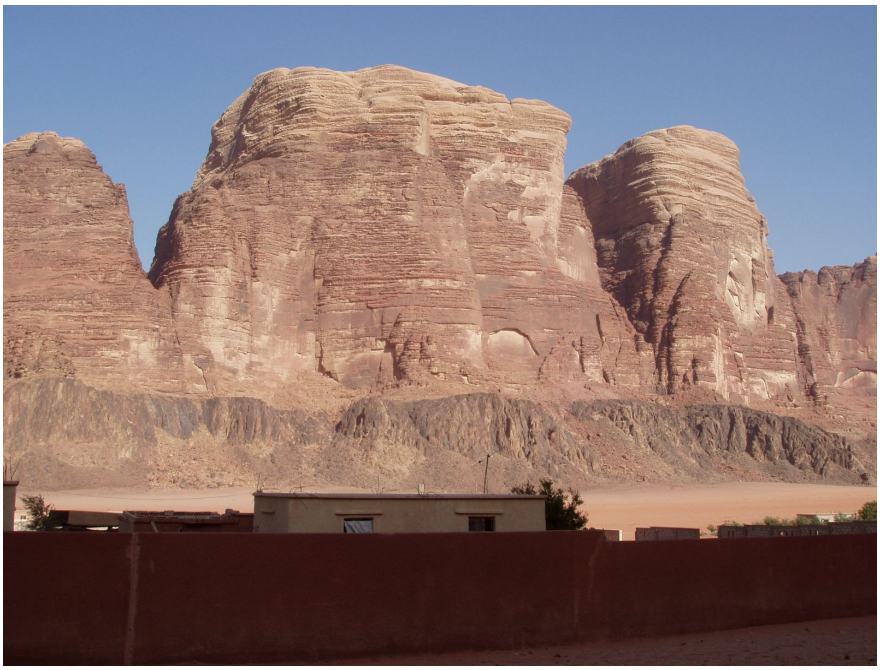


Output:









References:

[1] Content Based Image Retrieval Using Color Histogram A.Ramesh Kumar, D.Saravanan Sathyabama University, Chennai ,Tamil Nadu, India