Lab6: Object recognition. Bags of words

Esther Gonzalez and Wilmer Bandres

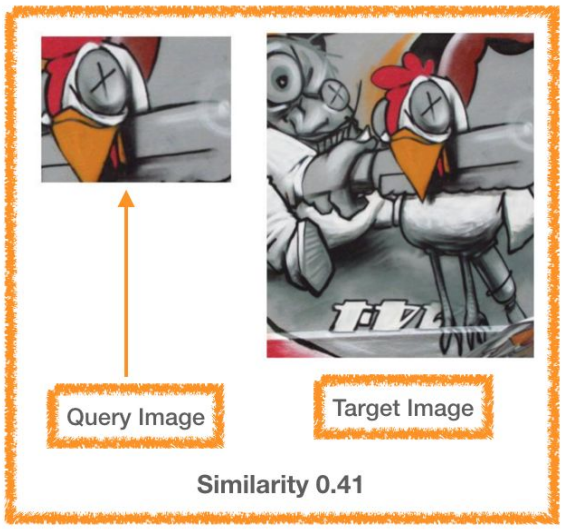
For this lab we implemented an object recognition system based on a text retrieval approach; generalized to nontextual information, bag of words. We have a database of images (from the data folder), and each image is represented by a set of SIFT descriptors, the descriptors are the actual database that will be used to look for a target image given a query image. The descriptors are used instead of the original image because using the descriptors the recognition can work despites any changes in the viewpoint, illumination and partial occlusion.

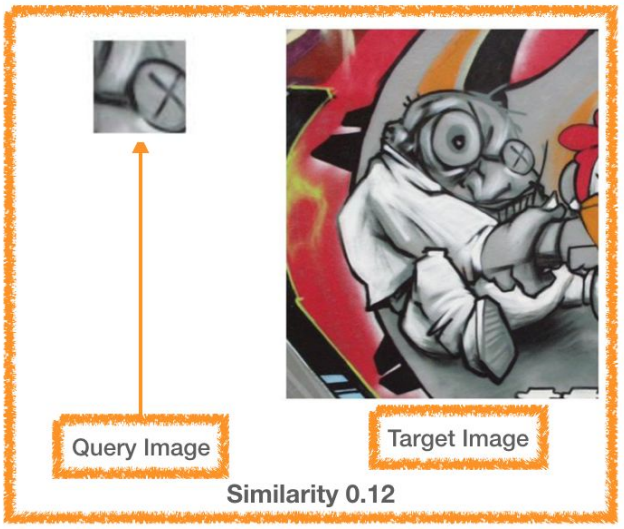
The descriptors are clusterized using k-means algorithm to improve the performance when querying the database. So, each descriptor is actually converted in one vector of K components based on a tf-idf approach (like the one in information retrieval), where *K* is the number of descriptors that are the centroids of the clusters in the descriptor space. For this implementation we first used few images from the database to check the performance, running kmeans with k=400, then we finally used all the images provided and k-means algorithm was run with a setting of k=2000.

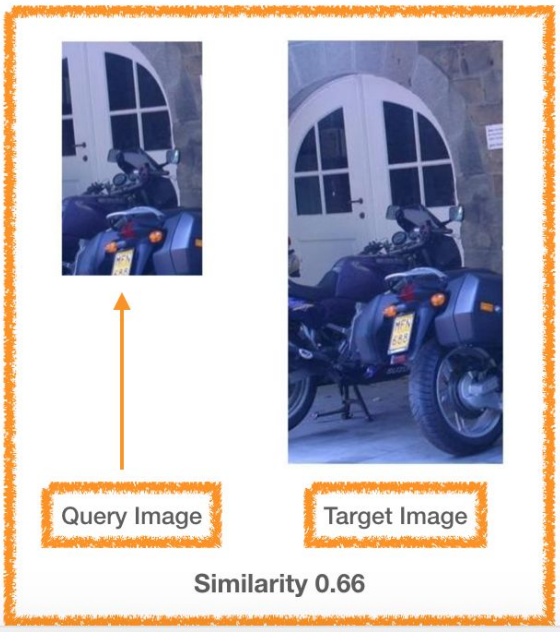
When a query is done, the descriptors of the query image are computed and converted to a vector of K components (as the images on the database). This new vector is not computed in the same way that the database images, instead of doing a clusterization we computed for each descriptor in the query image the nearest descriptor in the database which helps to compute the Nj and the mjn mentioned in the lab.

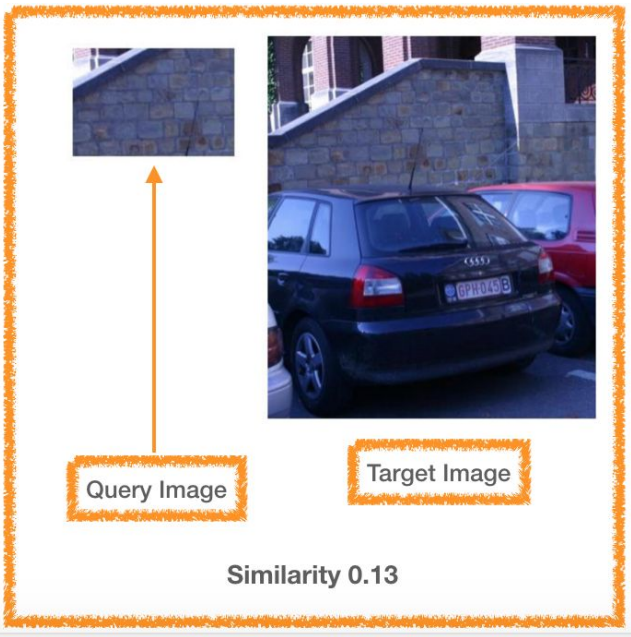
In order to have a good classification of the query image the number of cluster should be large enough, since a bigger number of cluster helps to differentiate between useful descriptors from the images in the database.

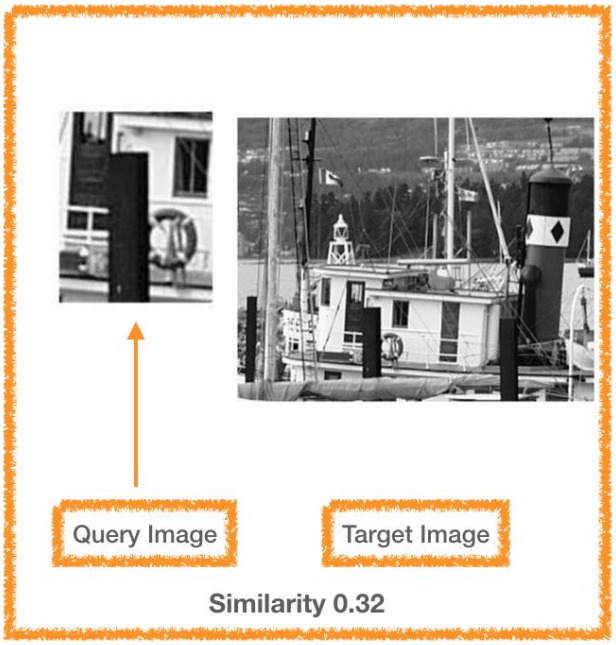
We ran the algorithm with several queries as shown below, the left images are the query images and the images in the right are the answer or the most similar image given by our algorithm. From the results we can see that the approach of doing visual search using a text-retrieval method works pretty well, and the use of SIFT descriptor allows finding target images that were taken from a different point of view (or that have a different illumination or are partially occluded or have a different size).

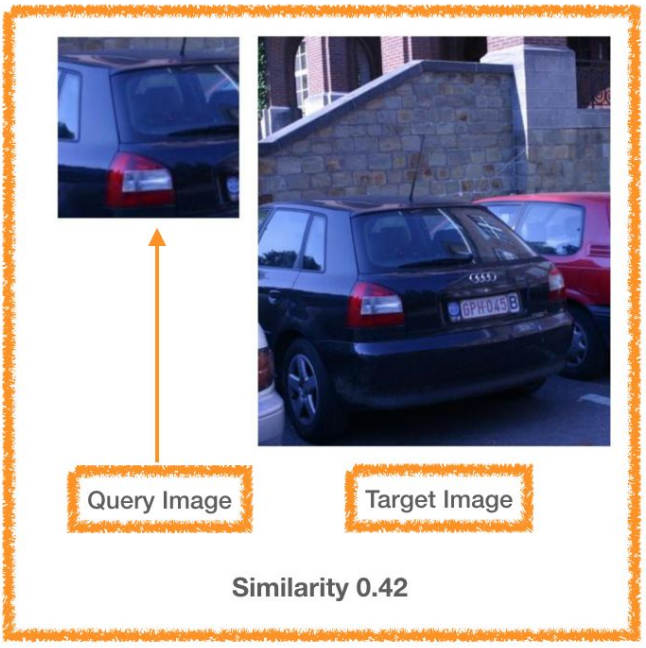












In order to run the script, just call the script similarity that is in the folder “Lowe_SIFT” which receives two parameters, the image (not the name of the file, but the image) and the number of clusters (when computing the bag of words).