Report Computer Vision Project 3

Single Michael
08-917-445

1 Video search with bags of visual words

For generating the data sets for this project I used two *mp4* movies. The first is a sequence of a *Breaking Bad* episode (called *breakingbad2*) and the second is a demo video found used for various purposes showing some kids playing with their toys (called *test_video*). These video files are stored in the folder *video/*.

I extracted from each movie their frames and stored them as png images in *frames/* using my Matlab function *extractVideoFrames*. For each frame converted to a grayscale image, I computed its corresponding SIFT data using VLFeat's function *vl_sift*. Hence, for each extracted frame I create a *.mat* file that contains

- the SIFT vectors (encoded in the rows)
- the name of the frame.
- the number of detected features
- the orientation of the patches
- the positions of the patch center.
- the scales of the patches.

You can generate your own mat files using my function *computeSiftDataOf*.

Last, I additionally assembled all mat files into one big mat file, which contains all mat files in a sequential order. This is just to simplify the coding that had to be done for the assignment. You can find all those mat files in the folder *data/*.

1.1 Raw Descriptor Matching

1.2 Visualizing The Vocabulary

1.3 Full Frame Queries

1.4 Region Queries