Computer Vision VU

Exercise Course: Assignments II

WS 2015/16

Harald Scheidl, 0725084
Thomas Pinetz, 1227026
Velitchko Filipov, 0726328

Assignment 4: Image Stitching

TODO

Assignment 5: Scene Recognition with Bag of Visual Words

Overview

First let's have a short look of how this classification method works.

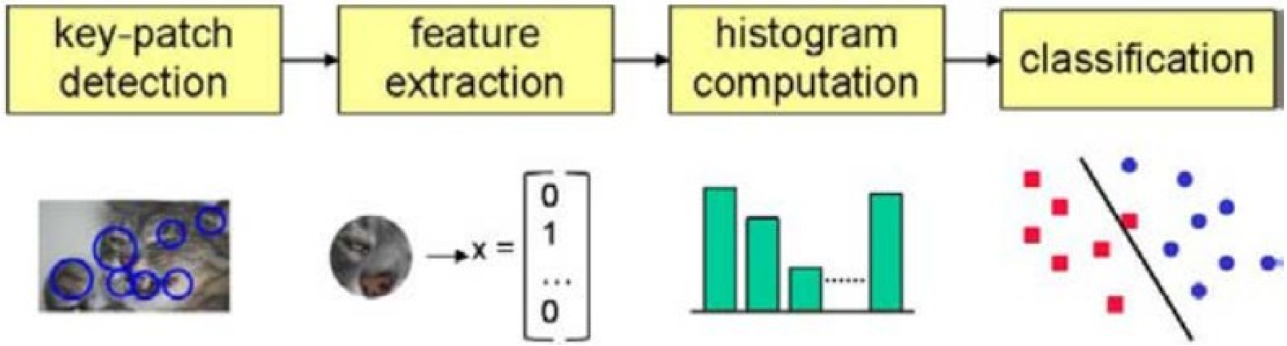


Fig: Overview of the classification step [Szeliski]

- **build a vocabulary of words**: a huge set of SIFT descriptors of all training images are collected and then clustered into 50 clusters in the 128-dim SIFT space. The centroids of those clusters are called visual words and are nothing but elements of a 128-dim vector space.
- **build training set**: now SIFT descriptors are calculated once more for the training images, and for each SIFT descriptor, the nearest visual word is searched by the KNN algorithm. For each image, a distribution (histogram) over the set of all visual words is calculated. This distribution is normalized to sum up to one. For each image, the distribution and the group label (e.g. “street”) are saved. One can again think of this distributions as elements of a 50-dim vector space.
- **classify**: for a new image, a number of SIFT descriptors is calculated and then assigned to the nearest visual word. Again, the distribution over all words is calculated. The KNN algorithm is used once more for classification, this time in the 50-dim space of distributions. The three (k=3) nearest elements of the training set are searched in this space and the class is determined via a majority voting.

Discussion of results

To measure the quality of this method, we use a confusion matrix C. Entries at the diagonal Cii represent the number of correct classified images with class label i.

Wrong classified images can be found on all non-diagonal entries Cij, i≠j: such entries represent the number of instances of (real) class i which are (wrong) classified as class j.

We added two visualizations of the confusion matrix to this report. In one visualization the entries are colour-coded, in the other visualization the numeric values can be found.

In both visualizations, rows represent the real classes, while the columns represent the classified classes.

The class labels are values between 1 and 8. The mapping between the numeric value of the classes and the names of the classes can be found in this list:

- 1 = bedroom
- 2 = forest
- 3 = kitchen
- 4 = livingroom
- 5 = mountain
- 6 = office
- 7 = store
- 8 = street

To get the percentage of correct classified images, we have to sum up all diagonal entries of C and divide this sum by the sum of all entries of C. This gives a value of 473/800=0.5913, which is about 60% as supposed in the assignment.

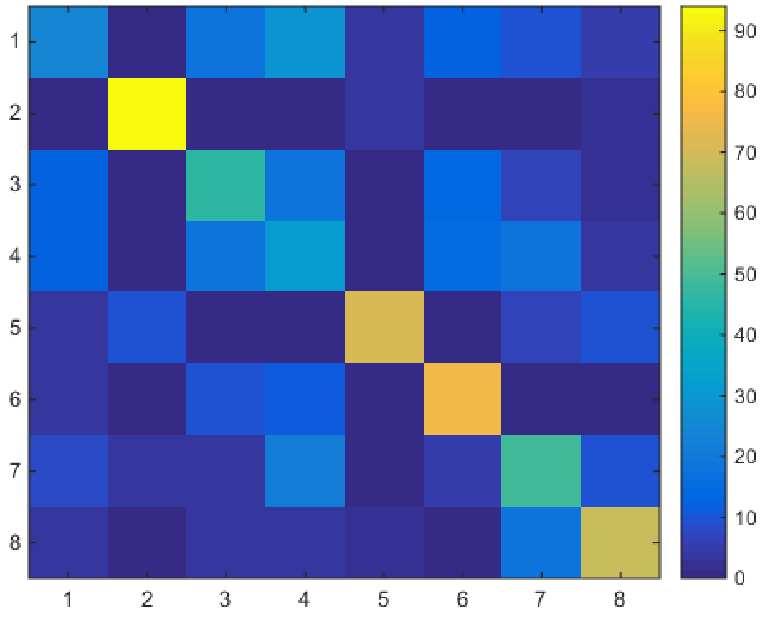
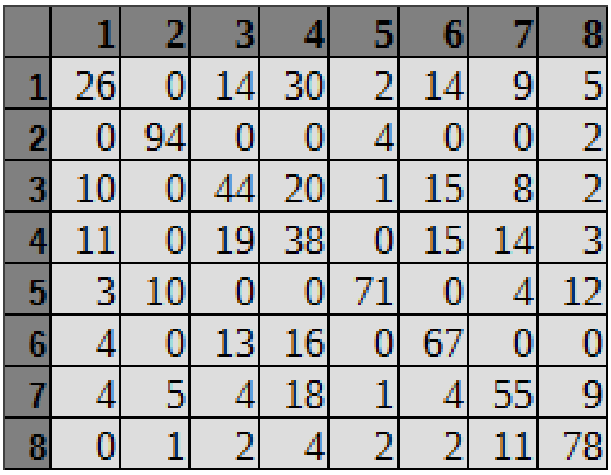


Fig: confusion matrix, colour-coded



	1	2	3	4	5	6	7	8
1	26	0	14	30	2	14	9	5
2	0	94	0	0	4	0	0	2
3	10	0	44	20	1	15	8	2
4	11	0	19	38	0	15	14	3
5	3	10	0	0	71	0	4	12
6	4	0	13	16	0	67	0	0
7	4	5	4	18	1	4	55	9
8	0	1	2	4	2	2	11	78

Fig: confusion matrix, numeric values

When comparing the numeric class labels with the class names, one can see that the classification of forests works pretty good. Also other “natural” (outdoor) scenes have a good classification rate.

Problems occur when classifying categories such as bedroom, kitchen or store. They appear to be pretty similar for this method, i.e. with respect to visual words.

Own images

We used three own images of three different categories. As one would expect after looking at the confusion matrix, there was no problem identifying the forest. Also the image of the mountain was classified correct.

The image of the street however was classified as a living room. Looking at the confusion matrix, one would expect a classification rate of about 78%. Inspecting the training images of the category street shows the reason for the incorrect classification: the training images represent streets in urban areas, while our own image shows the Overseas Highway which passes through non-urban areas.

Fig: forest (group=2), classified as forest (group=2)

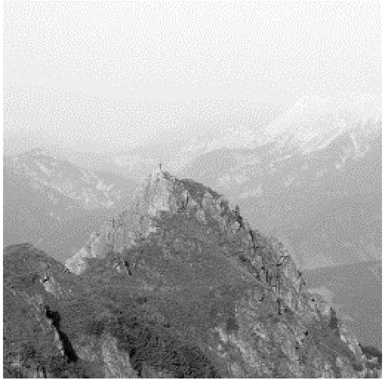


Fig: mountain (group=5), classified as mountain (group=5)

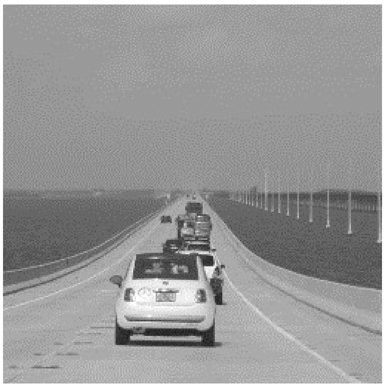


Fig: street (group=8), classified as living room (group=4)