Computer Vision 1, Master AI
Tutorial Lecture 4: Bag-of-Words

Thomas Mensink and Theo Gevers

1 Local Features

Consider the following image path (I):

$$I_{\mathrm{A}} = \begin{array}{|c|c|c|c|c|}\hline 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 0 \\ \hline 1 & 1 & 1 & 0 & 0 \\ \hline 1 & 1 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

1.a Compute the gradient G_x and G_y, using image filters. Which filters do you use?

1.b Compute the gradient magnitude

1.c Compute the gradient orientation (in degrees)

1.d Compute the HoG descriptor, using a 9 bin histogram

1.e Is the HoG descriptor invariant to overal lighting, *i.e.*, is the HoG descriptor of $I = 2 * I_A$ equal to the HoG descriptor of I_A?

2 Bag-of-Words

2.a Describe the basic steps of the Bag-of-Visual Words model?

2.b What is the difference between dense sampling en interest point sampling?

2.c Assume we have an image retrieval system, with 5000 images, 100 query images, and we use a BoW representation with 10K words, using SIFT descriptors. We observe that a BoW with *dense sampled* patches outperform BoW with *interest points* patches, when using precision@10 as evaluation measure. What could be the reason?

2.d Now we increase the dataset to 5M images, and interest points perform better. What could be the reason?

2.e Explain how k-Means clustering can be used to obtain a visual vocabulary.

3 Retrieval

3.a For retrieval, each image is described with 1000 interest points, each interest point is 128 dimensional. When finding matches between 2 images, how many computations are required?

3.b Comparing 1M interest points, takes 1 second. How long does it take to compare an image with a dataset containing 1M images?

3.c After retrieving results with BoW, we use geometrical verification to rerank the top 100 images. However, our evaluation shows no difference in precision and recall measured at k=100. Why?

3.d Explain why accuracy is not a good metric

3.e Compute Average Precision for the following relevance ranking: $[R, N, R, R, N]$, where R denotes relevant, and N not-relevant

4 Classification & Object Detection

4.a Explain how to train a "cat" vs "non-cat" classifier using linear classification (ie SVMs)

4.b Compute the number of box evaluations for a single image in a multi-class object detection problem

4.c Explain the idea of Selective Search?

4.d How does Selective Search increases the variations?