

Q and A for Web-Scale Image Retrieval

Sung-Eui Yoon
Dept. of Computer Science and Div. of Web Science and Technology
KAIST
sungeui@gmail.com
<http://sglab.kaist.ac.kr>

August 27, 2012

Abstract

In this document I have listed questions repeatedly asked from students taking the course, and reasonable answers to them. Note that answers may be incorrect, but are provided as a starting point to address questions.

1 Keypoint Localization

When we detect corners, we consider intensity of images. I think that color or other information of images should be also important. How can we consider them?

For example, Harris detect looks at gradients of intensity of images. We can naively apply the same concept (gradients) to colors of images. However, there may be a better way of considering colors or other information of images.

Why do we consider the auto-correlation equation, when we detect corners?

We would like to detect key points that are robust to many different configurations (see the lecture note) for image matching. Since we would like to extract key points independently in each image, we look at the autocorrelation function. For the case of corners, the autocorrelation function can have big changes, when we have small shifts (e.g., some u and v values in the lecture note) in a local window. In order to look at those changes in all the possible directions in an efficient manner, we apply the Taylor expansion with two orthogonal directions.

2 Scale-Invariant Region Detector

Why Laplacian-of-Gaussian (LoG) is better than the average intensity for the automatic scale detection?

LoG has been known to be an excellent detector for blobs as well as edges. Also, many Gaussian kernels have been known not to present any new information in coarser representations. Human visual responses can be modeled as Gaussian derivatives. Overall many studies suggest that LoG is one of the best functions that robustly detect the characteristic scale. For example, if an image has illumination change, the function of average intensity may not find good characteristic scales for images that shows the same object.

What are the differences between covariance and invariance?

Invariant features are ones that do not change even if there are illumination changes and so on. On the other hand, covariant features increases or decreases depending on other factors. We would like to design invariant or at least covariant features for various types of image retrieval.

3 Bag-of-Word Approach

We need to define the number of words for the bag-of-visual-words approach. Is there any automatic way of defining it?

As far as I know, there have not been many approaches on defining it in an automatic way. Many papers attempted different numbers of visual words, and measured the accuracy and the memory requirement as a function of the number of visual words. Then it chooses a particular number of words that gives a reasonable accuracy with a memory requirement for their problems.