Readings

- Grauman 2.1, 2.2, 3.3 Global & Local Representations, Local Descriptors
- Klette 9.2 Examples of Features
- Szeliski 4.1.2, 4.1.3 Feature descriptors, feature matching

- Online tutorial of SIFT (easy-reading, highly recommended)
- Comprehensive mathematical & algorithmic "dissection" of SIFT (article is very long, but is useful for understanding certain parts which may not be clear)
- Old Slides (from Prof. Kheng)

Summary

Motivation

- After finding the keypoints, we need to represent these locations (and their surroundings)
- Previous examples of feature descriptors: colour / texton histograms, mean gradients in x- and y-direction in a window, filter bank response, etc.

Descriptors

- Descriptors are vectors that mathematically characterize a region in the image
- Ideally, descriptors should be invariant / equivariant and do not change under image transformations, and also discriminative so we can identify and match corresponding points
- spatial histograms has some invariance to deformations, retains a rough spatial layout, e.g. in GIST descriptors
- when designing histogram-based features, need to pay attention to
  - what values are being histogrammed and number of bins
  - distance measure for comparing histograms
- orientation normalization: local descriptors can be fully rotation invariant if we first rotate the local patch according to a dominant orientation, e.g. in MOPS

SIFT Descriptor

- Scale-invariant feature transform is both a detector and a descriptor
- Keypoint detector based on multi-scale Laplacian of Gaussians, with addition of thresholding to discard keypoints from regions of low contrast and low-curvature (i.e. non-corners)
- Descriptor can be used stand-alone, e.g. w/ Harris corners; is a carefully engineered spatial histogram of local gradient orientations
- SIFT is invariant to scale and rotation, robust to changes in illumination and viewpoint

Feature Matching

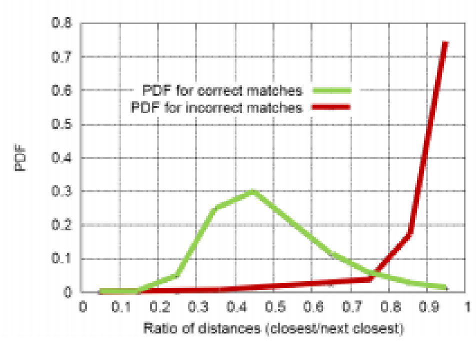
- Simply matching based on similarity may lead to ambiguous matches (e.g. fence posts)
- Better way (introduced by SIFT) is to use a ratio distance, need a high ratio between best match vs. next best match (ensures match is similar but next best match is not similar)
- Evaluate feature matches via an ROC (receiver operator characteristic) curve and AUC (area under the curve); higher is better, ensures high true positive rate and low false positive rate
- Matching is quadratic in complexity, can speed up with
  - indexing structures e.g. search trees, hash tables
  - fitting of transformation parameters (next week's lecture)

Demos

- Web demo of SIFT to try out different parameters

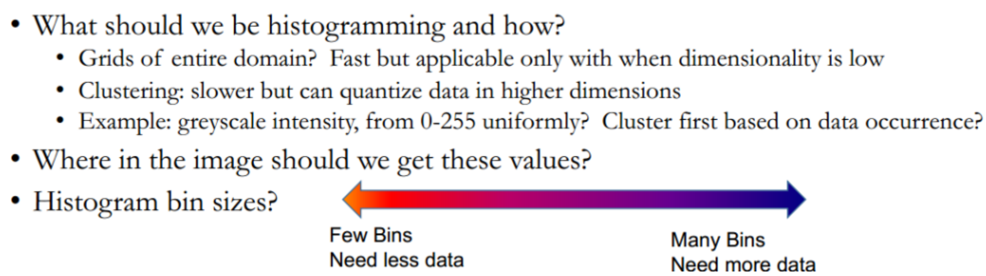
FAQs

Q: For the graph on slide 29, why does the curve for correct matches go down after a certain ratio/threshold? Wouldn't any previously marked positive features still be positive when the threshold increases, which should result in the same number of correct matches?



A: On the x-axis is the **ratio** of distances, i.e. ratio of best match / next best match. If you take a v. high ratio, then it means that your best match and next-best match are very similar. So you end up with lower number of correct matches likely because you include a lot of ambiguous matches (similar to our fence-post example).

Q: What do grids of entire domain mean? So we separate the entire image into grids and histogram each grid separately, e.g.like spatial histograms? What about clustering? Is this the same ideas as textons? (like clustering our filter bank responses in a particular region?)



A: "grids of entire domain": Suppose we are binning greyscale intensity. One option is to divide bins from [0 255] (entire domain) and distribute the bins equally (e.g. 16 bins, each bin covers range of 16 greyscale values). Alternatively you can check image statistics, e.g. 90% of the data is concentrated in greyscale of 100-200, so we keep bulk of bins in this area and make them more fine-grained, and allocate larger bins where not much data is distributed.

Q: In MOPS, what does it mean to rotate to horizontal? If my dominant gradient is 30 degrees, then I make the 30 degrees the new 0 degrees/origin? What is the difference between rotate to horizontal and orientation normalization?

A: rotate to horizontal is the orientation normalization. to picture -- the slide does this. the same patch on the mountain gets rotated to horizontal.

Q: For the GIST descriptor, does the image "patch" mean a window around a keypoint?

A: No. The patch is whichever part of image you want to compute the descriptor on. Usually for GIST, this is the entire image, but you can also compute GIST on 1/4 of the image (or some other large portion of the image)