Class 19 — Initialization Based on Matching Keypoints

Overview

This document provides an overview to using "interest points" or "keypoints" for initialization. The discussion is based on three papers:

- D.G. Lowe. Distinctive image features from scale-invariant keypoints. to appear in *International Journal of Computer Vision*. This gives the clearest description and most effective keypoint detection technique I know of. It is the most important to read for this class.

- M. Brown and D.G. Lowe. Recognizing panoramas. in *International Conference on Computer Vision*, 2003. This shows an application of keypoints. We'll come back to this in a later class.

- C.V. Stewart. Robust parameter estimation in computer vision. *SIAM Reviews* **41:3**, 1999. This describes robust estimation in the context of computer vision problems.

In Lecture 18, we discussed a variety of initialization methods, including manual initialization, sampling of parameter space, and alignment based on moments. We also discussed the fact that initialization is still important even though we have a variety of methods at our disposal to increase the domain of convergence of registration. Today we'll discuss use of "keypoints" or "interest points". In some contexts, preliminary matching of keypoints or interest points is so powerful that no further refinement of the estimated transformation is necessary. In general, however, refinement of the estimated transformation is still necessary following generation of the initial estimate.

Summary of Main Algorithmic Steps

1. Generating/detecting keypoints: we'll focus on Lowe's generic technique, but be aware that application-specific methods, such as detection of vascular branch points, are also used.

2. Describing keypoints in a manner that is "invariant" (or nearly invariant) to the unknown transformation.

3. Matching of keypoints to generate an initial transformation

Detecting Keypoints — Goals in Developing a Generic Method

- Distinctiveness
- Scale independence
- Repeatability.

Steps to Detection

1. Formation of scale space (see Fig 1):
   (a) Gaussian pyramid
   (b) Laplacian of Gaussian,
   (c) Difference of Gaussians

   Aside: Why does this do what we want?

   - Think of detecting an isolated spot in an image: the peak response will be at a scale where the positive region of the Laplacian of Gaussian matches the width of the spot.

2. Peaks (valleys) must be found simultaneously in x, y, and σ (see Fig 2).

Be sure to take a careful look at the extensive empirical analysis leading to the choice of σ.

Keypoint Description - Goals

- Invariant to translation, rotation, scale and illumination
- Nearly invariant to affine distortions

Aside — What is An Invariant?

- An invariant is a quantity that does not change with the application of any element from a group of transformations
- Under translations: lengths and orientations are invariant
- Under rigid transformations: vessel widths, lengths, differences between angles are all invariant

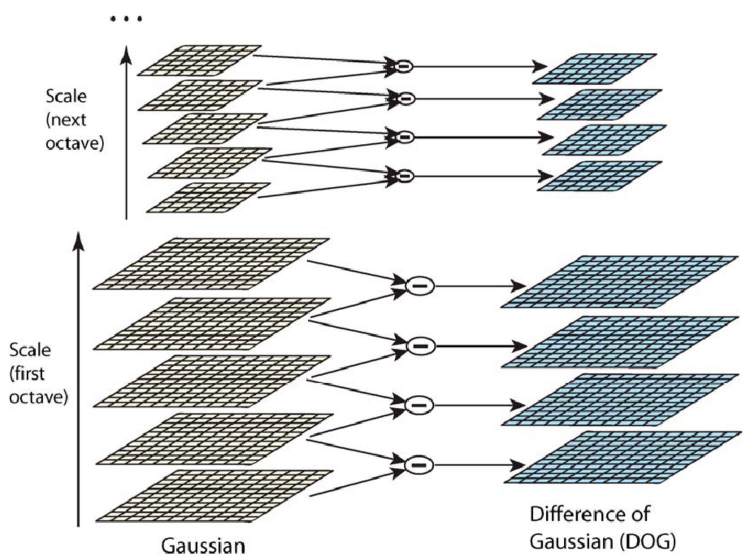

Figure 1: The left is the set of scale space images. The right is a set of difference-of-Gaussian (DoG) images. After each octave, the Gaussian image is down-sampled by a factor of 2. and the process repeated.

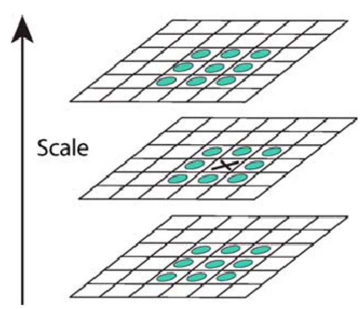

Figure 2: Max and min of the DoG images are detected by comparing a pixel (Marked as X) to its 26 neighbors in 3x3 regions at the current and adjacent scales (marked with circles).

- Under similarity transformations: ratios of lengths and differences in angles are invariant

- Under affine transformations: ratios of lengths along a single line segment are invariant.

- In general, transformation groups with more degrees of freedom have fewer invariants.

Invariance is important for describing a keypoint in a way that allows it to be "recognized" independent of camera viewpoint.

Keypoint Description

- Find the dominant orientation at the selected scale. Everything else is described relative to this dominant orientation.

- Divide up image region surrounding keypoint location into bins (4×4) and compute a gradient-magnitude-weighted histogram of (8) orientations in each.

- This uses linear interpolation to split each point's contribution among different bins. It also Gaussian weighting to downgrade the contribution of points further (relatively) from the keypoint.

- Collect the $4 \times 4 \times 8$ histogram values into a 128-component vector and normalize to unit magnitude. The normalization removes linear changes in illumination.

- Mathematically the result is only similarity invariant, not affine invariant.

Aside(1): Description of Landmarks in the Retina

- Ratios of widths of vessels that meet to form the landmark

- Differences in orientations of vessels that meet

  - Sometimes we use just the orientations themselves because the transformations do not have a great deal of rotation

Aside(2): Computation of Dominant Orientation

- It is important to fix the dominant orientation first.
- For Lowe keypoints, the following is done for each selected keypoint with the chosen Gaussian smoothed image.:
  - compute the gradient orientation and magnitude for each (x,y) in the circular neighborhood of the keypoint.
  - A orientation histogram is formed, with each sample weighted by its graduate magnitude and by a Gaussian window (See Fig7 in Lowe's paper).
  - Peaks in the orientation histogram correspond to dominant directions of local gradients.
  - If multiple peaks, create multiple keypoints (same location and scale, but different orientations).
- How about the landmarks in the Retina?

Matching of Keypoints

- Form list of keypoints and their descriptor vectors in each image
- For each keypoint
  1. Find the closest and the second closest descriptor vector in the other image. Compute the keypoint distance for each.
     - Representation and search are based on a modified k-d tree
  2. Form ratio of descriptor distances.
  3. Retain the closest as a match if the ratio is less than 0.8. Correct matches usually have a much smaller ratio than incorrect matches.

Initial matching in the Dual-Bootstrap ICP registration algorithm is MUCH different: see Lecture 16 for details.

Determining Which Matches Are Correct and Estimating the Transformation Parameters

Problem:

- Given are the keypoint matches, $\mathcal{C} = \{(\mathbf{g}_k, \mathbf{f}_k)\}$, with at most one per keypoint, $\mathbf{g}_k$.
- Many of these matches, perhaps more than half(!), are incorrect.
- We also know, approximately, how much error there is in the positions of the keypoints: the standard deviation approximately matches the scale of the keypoint (in scale space).

What we are going to discuss is not the technique used by Lowe. Instead, we'll describe a method that has been used in a large number of other applications: see *SIAM Review* paper for examples as of 1998.

RANSAC / MSAC: Intuition

- We need to use robust estimation. In other words, we want to find the parameters that minimize

  $$E(\mathbf{\Theta}) = \sum_k \rho(\|\mathbf{T}(\mathbf{g}_k; \mathbf{\Theta}) - \mathbf{f}_k\| / \sigma_k)$$

  (we use Euclidean distances for keypoint locations)
- Problem: we can't start with a gradient-descent technique such IRLS — we don't know where to start.
- Solution:
  - Generate and test transformation estimates from random (minimal) subsets of the correspondences.
  - If "good" correspondences are chosen, then other correct correspondences will have small alignment error distances.

RANSAC / MSAC: Procedure

The following procedural description is specific to estimating 2d affine transformations:

1. Repeat:
   (a) Choose 3 correspondences at random from $\mathcal{C}$.
   (b) Generate an affine transformation from the chosen 3 correspondences. Let $\hat{\mathbf{\Theta}}$ be the parameters of the estimated transformation.

(c) Evaluate $E(\hat{\mathbf{\Theta}})$, and retain $\hat{\mathbf{\Theta}}$ as the current estimate if $E(\hat{\mathbf{\Theta}})$ is the smallest thus far.

Until S sets of 3 correspondences (S transformations) have been tested.

2. Gather "inliers" to the best estimate and, if desired, refine the estimate based by applying a least-squares transformation estimator using only these inliers

Notes

- The objective function is no different from the objective functions we've already explored with M-estimators and IRLS. This is is really just a different form of an M-estimator search.
- The original technique, RANSAC, did not use a smooth ρ function.
- The number of sample sets, S, can easily be derived based on probabilistic considerations (see pg 519 of *SIAM Review* paper).

Other Options

- Other robust estimators such as "least-median of squares" and its variants.
- Hough transforms
  - See Lowe paper for how this is used in object recognition

Summary

- Importance of initialization and keypoint / interest point based initialization.
- Keypoint detection
- Invariant keypoint description and matching
- Estimation and outlier removal using random sampling search.

Looking Ahead to Class 20 on Mutual Information (MI)

We'll discuss the two papers published nearly simultaneously that started the push toward use of MI in medical image analysis. These are Maes, et al. from Leuven (Belgium) [1] and Wells, et al. from MIT / Brigham and Women's [3]. A third paper, by Studholme, et al. [2] contains a good motivation for the idea of mutual information and introduces "normalized mutual information".

References

[1] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging.*, 16(2):187–198, 1997.

[2] C. Studholme, D. Hill, and D. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Patt. Recog.*, 32:71–86, 1999.

[3] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multimodal volume registration by maximization of mutual information. *Med. Image Anal.*, 1(1):35–51, 1996.