

ENEE631 Project 2 Proposal - Extracting Image Features using Scale Invariant Feature Transform (SIFT)

Naotoshi Seo, David A. Schug

April 23, 2007

1 Introduction

Image matching is a fundamental aspect of many problems in computer vision, including object or scene recognition, solving for 3D structure from multiple images, stereo correspondence, and motion tracking. Scale-invariant feature transform (or SIFT) proposed by David Lowe in 2003 [1] is an algorithm for extracting distinctive features from images that can be used to perform reliable matching between different views of an object or scene. The features are invariant to image scale, rotation, and partially invariant (i.e. robust) to change in 3D viewpoint, addition of noise, and change in illumination. They are well localized in both the spatial and frequency domains, reducing the probability of disruption by occlusion, clutter, or noise. Large numbers of features can be extracted from typical images with efficient algorithms. In addition, the features are highly distinctive, which allows a single feature to be correctly matched with high probability against a large database of features, providing a basis for object and scene recognition.

2 Methodology

SIFT

The major steps in the computation of SIFT are [1]

1. **Scale-space extrema detection** - a specific type of blob detection where each pixel in the images is compared to its 8 neighbors and the 9 pixels each (corresponding pixel+8 neighbors) of the other pictures in the series.
2. **Keypoint localization** - keypoints are chosen from the extrema in scale space.
3. **Orientation assignment** - for each keypoint, in a 16x16 window, histograms of gradient directions are computed (using bilinear interpolation).
4. **Keypoint descriptor** - representation in a 128-dimensional vector.

Keypoint matching

SIFT detects feature points in each image individually, thus we must find corresponding pairs in two images for uses in object recognition, structure from motion, stereo correspondence, and motion tracking. [1] also described how to do it.

2.1 Individual Task Lists

We separated individual tasks as follows so that individual workloads are almost equivalent:

Naotoshi Seo

1. Scale-space extrema detection
 1. Gaussian and Difference-of-Gaussian (DoG) filter
 2. Generate Gaussian and DoG pyramids
2. Keypoint localization
 1. Detect local extrema in the DoG pyramids
 2. Remove low contrast extrema
 3. Remove points which is matching with edge points

David A. Schug

3. Orientation assignment
 1. gradient orientations in regions surrounding each keypoint.
4. Keypoint descriptor
 1. representation in a 128-dimensional vector

Keypoint matching

1. In the sense of the nearest-neighbour

3 Extra Plans

We will work for some additional tasks if time allows.

3.1 Panoramic Image Stitching

Panoramic image mosaicing has an extensive research literature and there are several commercial offerings that come bundled with today's digital camera. Matthew Brown and David Lowe [3] presented an application of SIFT to fully automatic construction of panoramas in 2003. Panoramic image mosaicing has following steps:

1. Extract SIFT features - for all n images
2. Feature Matching - by k nearest-beighbours
3. Image Matching - by RANSAC
4. Bundle Adjustment
5. Multi-band blending.

3.2 Structure from Motion

The structure from motion - recovering scene geometry and camera motion from a sequence of images - is an important task and has wide applicability in many tasks, such as navigation and robot manipulation. We will use a factorization method developed by Tomasi and Kanade [6] to recover shape and motion under an orthographic projection model. This method requires to know tracking points in the image sequence, thus we will use SIFT as a tracking method for the purpose.

The factorization method is already implemented by the author, N. Seo [8], as a course project in ENEE731, University of Maryland, 2006.

Bibliography

- [1] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 60, 2 (2004), pp. 91-110.
- [2] P. J. Burt and E. H. Adelson. "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, COM-31(4):532-540, 1983.

For Application to Panoramic Image Sticking

- [3] M. Brown and D. Lowe. "Recognising panoramas," *In Proceedings of IEEE International Conference on Computer Vision*. IEEE Computer Society, 2003. 64
- [4] H. Shum and R. Szeliski, "Panoramic Image Mosaics," Tech. Rep. MSR-TR-97-23, Microsoft Research, 1997.
- [5] P. J. Burt and E. H. Adelson. "A multiresolution spline with application to image mosaics," *ACM Transactions on Graphics*, 2(4):217-236, 1983.

For Application to Structure from Motion

- [6] C. Tomasi and T. Kanade, "Shape and motion from image streams – a factorization method," *International Journal of Computer Vision*, 9(2):137–154, 1992.
- [7] CMU VASC Image Database: hotel <http://vasc.ri.cmu.edu/idb/html/motion/hotel/index.html>
- [8] Naotoshi Seo, "ENEE731 Project - Shape and motion from image streams: a factorization method," 2006. <http://note.sonots.com/?SciSoftware%2FFactorization>