As a part of the course Advanced Image Analysis at the Technical University of Denmark (ITU), we conducted a mini-project where we used SIFT (Scale-Invariant Feature Transform) (Lowe, 2004) to identify initial corresponding points between two views of the same scene. Knowing the geometric setting of the problem, we were able to use RANSAC (Kovesi, 2000) to determine the set of inliers and to estimate the transformation between the images. The framework allowed us to implement automatic image stitching and automatic estimation of fundamental matrix for stereo view.

SIFT produces a set of localized corresponding points, which are further RANSAC filtered for inliers. One of the images is then warped according to the homography (two point correspondence transformation) and images are stitched together using the “hat” weighting function when stitching multiple panoramas.

For image matching, descriptor vectors of all keypoints are computed. The components of the SIFT framework for keypoint detection are as follows:

1. Scale-space extrema detection. Using a cascade filtering approach a set of candidate keypoints are selected by taking the extrema of all Gaussian image sequences across the range of scales. Local maxima and minima are then detected over all scales and image locations. The SIFT approach generates a large number of features, directly covering the range of scales and rotations.

2. Keypoint localization. Each candidate keypoint is fit to a 2D Gaussian model to estimate location and scale. The points with variance greater than a predefined threshold and partly isolated edge points are rejected.

3. Orientation assignment. Based on local image gradients, each keypoint is assigned a direction. In case of more strong directions, additional keypoints are created.

4. Keypoint descriptor. This is accomplished by sampling image gradient magnitudes and orientations around each keypoint and dividing those in an array of orientation histograms into bins. SIFT produces a large number of features, directly covering the range of scales and rotations. The entries of all histograms are then put in a descriptor vector which is normalized to reduce the effects of illumination changes.

The number of matching keypoints falls drastically, but matches are still mostly correct. After 15 iterations, the drop in the number of matches slows, but false matches represent up to one quarter of all matches.

References:


APPLICATION — USING SIFT

Scale-Invariant Feature Transform (SIFT): Performance and Application

Vedrana Andersen  
The IT University of Copenhagen  
Email: vedrana@itu.dk

Lars Pellarin  
The IT University of Copenhagen  
Email: pellarin@itu.dk

Renée Andersen  
The IT University of Copenhagen  
Email: renee@itu.dk

We tested SIFT’s performance in a series of controlled tests. A testing image was matched against itself, yet modified by various transformations. Knowing the original transformation, we were able to get true or false matches. A match was labeled false if the matched keypoint was at a distance of more than 2 pixels. We tested scale and rotation invariance, robustness to perspective transformation and gave the presence of noise. The number of matches found was compared to the reference.

For a large number of keypoints matched for each scale we tested, highest number of matches, at surprisingly, for the 1:1 scale, but in general number of matches is not changing dramatically for other scales. Number of false matches varies around 2%.

We also tested SIFT’s invariance to rotation. The set of candidate keypoints matching the target image were rotated 10 degrees every 10 iterations, approximately. 300 keypoints matched; false matches shown in red. For a large number of keypoints matched for each scale we tested, highest number of matches, at surprisingly, for the 1:1 scale, but in general number of matches is not changing dramatically for other scales. Number of false matches varies around 2%.

SIFT’s robustness was tested under various perspective transformations: the testing image was shrunk and one of it’s sides scaled by factors from 1 to 0.5. Shown here is a projectional shear of factor 0.5.

The number of matching keypoints falls steadily with decreasing or partly fixing the area of the warped image falls with ½. Results included for partners in the warp case 1½, increase 1½.

We also tested SIFT’s invariance given the presence of noise. The image is modified with noise, we are able to test SIFT with different levels of noise with each iteration. After the first additions of noise the number of keypoints matched drops slowly, but matches are still exactly correct. After the second 5 iterations noise is added, but false matches represent up to one quarter of all matches.

Scale:

Rotation:

Projection:

Noise:

First row: SIFT keypoints for two different images of the same scene. Keypoints are displayed in red and matched keypoints in yellow. First column: SIFT keypoints for the left image and SIFT keypoints for the right image. Second column: SIFT keypoints for the left image, having the setting of the scene; no two points are false so the RANSAC and find the outliers. Bottom set of images: SIFT found 927 and 1176 keypoints in the left and right image, respectively, which resulted in 278 matches. Applying RANSAC this resulted in 278 outliers. At the right of the figure are original images with the set of corresponding points and correspondences marked with line. Bottom set of images: SIFT found 927 and 1176 keypoints in the left and right image, respectively, which resulted in 278 matches. RANSAC estimated only 3 outliers, having a set of 18 matches.

The fundamental matrix is essential for stereo-view geometry; it describes mapping between points in one image and corresponding points in another image. At least eight point correspondences are needed to estimate the fundamental matrix. As in the previous example, SIFT produces the initial corresponding points, and RANSAC is then used to fit the fundamental matrix. The number of matches returned by SIFT varies depending on the size of the overlapping area, but the percentage of outliers is always large enough for RANSAC to estimate a homography that results in a satisfactory stitching.

Top row of images: Stitching the two images of the EU building. Original images. SIFT detects (with sound clamping) RANSAC (orange curve), and final stitch. Size of final image: 1024 x 768 pixels. Below the final image:

Bottom row of images: Another image mosaic, with only 96 keypoint matches and 18 outliers. Size of final image: 696 x 1174 pixels.