EE838 Assignment 2:
Evaluation criteria of local descriptors

Jonas Nikula
20176392

2017/09/20

Answers to presented questions

A. Briefly explain about 'homography (plane projective transformation)', and 'homography matrix'. What is the DoF (degree of freedom) of a homography matrix?

Homography maps points in the plane of the real world to points on the image plane. The homography matrix describes this mapping, and it can be used to translate world coordinates to image coordinates. This can be done by simply multiplying the image coordinates with the matrix.

The degrees of freedom of a homography matrix is given by the count of elements that are unknown, or free to change. The homography matrix has the form 3×3, and one of those elements is the scale constant 1, so we have $3 \times 3 - 1 = 8$ degrees of freedom.

B. What is the difference between 'region detectors' and 'region descriptors'?

A region detector detects simple features in an image. These simple features are often edges or corners, places where the intensity gradient is large. A region descriptor takes these detected features, and tries to group them. The purpose of region descriptors is to make the feature detection more robust to scale, orientation and intensity changes.

C. What is the difference between 'Harris–Laplace' and 'Harris–affine' region detectors? And, what is the difference between 'Harris–' and 'Hessian–' region detectors?

'Harris-Laplace' is a scale invariant corner detector, 'Harris-affine' is an affine invariant detector. 'Harris-' detectors detect corners, 'Hessian-' detectors detect blobs.

D. Write the definition of 'repeatability rate', which is defined in the paper [1]. Briefly explain about the method of measuring a repeatability rate from a pair of images.

The 'repeatability rate´ is the percentage of detected points simultaneously present in two images. The repeatability rate is of course the better the higher it is, since more correctly detected shared features makes for better matching.

The repeatability rate can be calculated by dividing the number of points repeated in both images by the total number of detected points.

E. What is the ground truth for measuring 'recall' and '1−precision' in [1]?

The ground truth is gotten by manually selecting corresponding regions from different images of the same scene.

F. Write the mathematical expressions of calculating 'recall', and '1−precision' in [1], along with brief explanations.

The mathematical expression for calculating recall is shown in 1. #correspondences is the number of manually selected points that are present in both images. #correct matches is the number of correspondenses made by the descriptors that match the manually selected correspondences.

$$recall = \frac{\#correct\ matches}{\#correspondences} \tag{1}$$

The mathematical expression for calculating 1-precision is shown in 2. The #correct matches variable is the amount of detected matches that are the same as the ground truth, and #false matches are the matches that aren't the same.

$$1 - precision = \frac{\#false\ matches}{\#correct\ matches + \#false\ matches} \tag{2}$$

G. There are three typical matching strategies for descriptor matching (4.1.1 in [1]). Explain each matching strategy, i.e., threshold−based, nearest neighbor−based, and distance ratio of nearest neighbors.

With a **threshold−based** matching strategy, the regions are matched if the distance of their descriptors is below a certain threshold. A region can have several matches.

A **nearest neighbor−based** matching strategy matches the regions which are closest to each other, as long as the distance between is also below a set threshold. Using this strategy a region only has one match.

In a **distance ratio of nearest neighbors** strategy the closest descriptors are also matched, but only if the ratio of the distance to the closest descriptor divided by the distance to the second closest descriptor is below a threshold.

H. Briefly explain about the dataset for evaluation, which is provided in [1] with six types.

The dataset contains six types of image pairs:

1. Rotated pairs, where the other image is a rotated version of the first one.
2. Rotated and zoomed pairs, where the zoom level also differs.
3. Pairs where the viewpoint changes, ie. the other image is taken from another position.
4. Pairs where the other image is blurred.
5. JPEG compression on one image
6. Light change pairs, ie. pairs where one image has a lower intensity than the other.

The images are also of different types of scenes; Structured scenes with clear objects and regions, and scenes of repeated textures. There's always a homography between the images, because they are either taken of the same planar scene, or the camera position does not vary.

Questions about the code

1. Explain about the data in the file 'img1.haraff.sift'. What do the first two lines indicate? For each line after the 3rd one, what are the components?

The first two lines describe the size of the data in the file. The first line is the number of columns, the second the number of rows.

Each line after the 3rd one is a descriptor for a region. The first two components are the center x and y coordinates of the ellipse. The following three components are the ellipse parameters that define the region. The rest are descriptor invariants.

2. Analyze the differences among the three matching strategies based on the experimental results.

(This analysis is based on the results of the matlab code)

Both nearest neighbor based strategies had 998 ground truth matches. The nearest neighbor based strategies functioned quite similarly, they had the exact same number of correct matches, 756. However, the distance ratio thresholded nn strategy had fewer total matches, 1758 compared to 2050 of the normal nn. The distance ratio nn has a recall rate of 75.8%, which is the same for the normal nn. The distance ratio nn has a 1-precision rate of 57.0%, and the normal nn has a rate of 63.1%.

The threshold based strategy had 3685 ground truth matches. It had a total of 74711 matches, of which 2339 were correct. This means it had a recall rate of 63.5%, and a 1-precision rate of 96.9%.

In conclusion, the distance ratio nearest neighbor strategy seems to be the best, with the same recall rate as the normal nn, but with a better 1-precision rate. The simple threshold based strategy is worse than both the nearest neighbor strategies.