Repeatability Performance Evaluation of Local Feature Detectors on Pyramid Images*

Yunsheng Zhang*, Zhengrong Zou

School of Geosciences and Info-Physics, Central South University, Changsha 410083, China

Abstract

To generate corresponding points for large size satellite, aerial or high resolution close-range image, it is always needed to downscale the original image to low resolution image via pyramid image generating to detect local features. So this paper studies the affection of the pyramid image on the repeatability performance of the state-of-the-art local feature detectors. For generating pyramid image with ground truth data, a method for calculating the homography parameters between pyramid images of stereo pair with homography parameters is proposed. Benchmark data sets with ground truth are used to generate pyramid images and derive homography parameters directly. Experiment results revealed that pyramid image can reduce the affection of scale change and image distortion to some extent.

Keywords: Local Feature; Performance Evaluation; Pyramid Image; Homography Parameters

1 Introduction

Local feature detection is important for image matching, image orientation, image registration, 3D reconstruction, object recognition and other applications. Recently, with the development of SIFT algorithm, several scale, rotation and affine invariant local feature detectors are proposed in computer vision community [1, 2, 3]. Tuytelaars and Mikolajczyk conclude that there are many properties for idea local feature, such as repeatability, distinctiveness, accuracy, locality, quantity and efficiency, while repeatability is most important among them in many applications [3]. Due to their performances, researchers from Photogrammetry community use these feature detectors to extract tie point for aerial triangulation or control points for image registration [4, 5]. But the detectors are commonly developed for low resolution images, video images or small size images, and can’t be directly used to large size digital aerial images, satellite images or high resolution close range images. These high resolution images are often downscaled to adapt to above local

*The work described in the paper is supported by China Postdoctoral Science Foundation funded project (No. 74141000326), Fundamental Research Funds for the Central Universities (No. 2012QNZT078), National High Technology Research and Development Program of China (No. 2012AA120801), National Basic Research Program of China (No. 2012CB719903).

*Corresponding author.
Email address: zhangys@csu.edu.cn (Yunsheng Zhang).
feature detectors. However, what the affections of the downscale processing on the performance of local feature detectors are not known to us.

Performance evaluation work is an important task in developing local feature detector. Several methods and criteria have been proposed to evaluate the performance of local feature detector. Repeatability rate are presented to evaluate the possibility of one feature re-appear in other images [6], and it is become famous for local feature performance evaluation. In the work of Cordelia Schmid et al., some stereo data with homography transformation parameters as ground truth data are used to evaluate the performance of the state-of-the-art local feature detector, and Harris is found to give the best performance [6]. For complex 3D object scene, epiploar line intersection in a triplet image set is used as ground truth data [7]. Several local feature detectors take scale into consideration, but the performance of the detector on different resolution images are not studied thoroughly. So this paper uses some benchmark data set to derive different resolution images to evaluate the performance of the state-of-the-art local feature detectors.

The rest of the paper is organized as follows. Section 2 reviews the state-of-the-art local feature detectors, and introduces five common used detectors which are evaluated in this paper. Section 3 focuses on the performance evaluation criterion and proposes a method to calculate homography transformation parameters for pyramid images from original image pair. Experiment data and the results are discussed in Section 4. Finally, some conclusion and remarks are given.

2 State-of-the-art Local Feature Detectors

During last decade, several local feature detectors have been proposed. Tuytelaars and Mikolajczyk gives a survey on the state-of-the-art local feature detector, they generally categorize local feature detector to three classes: The first kind is corner detector which often extracts corners from image. The corners locate where the brightness gradient of the image is “rich” in a small neighbourhood. The corners can be junction points, interest points or edges points with high curvature. The second is region detector and the third is blob detector. Five common used local feature detectors which are evaluated in this paper are introduced as following sections.

2.1 FAST

FAST detector is proposed by Rosten and Drummond for corner detection [8]. It compares the gray value of one point to its circle neighborhoods, and then a point is classified as a corner if a sufficiently large set of pixels on a circle of fixed radius around the point such that these pixels are all significantly brighter (resp. darker) than the central point. After that, a non-maxima suppression processing is carried out to obtain final interest point result.

2.2 Harris

Harris detector is proposed by Harris and Stephens [9]. It depends on an auto-correlation matrix defined as follows:

$$M = \sigma_D^2g(\sigma_I) * \begin{bmatrix} I_x^2(x, \sigma_D) & I_x(x, \sigma_D)I_y(x, \sigma_D) \\ I_x(x, \sigma_D)I_y(x, \sigma_D) & I_y^2(x, \sigma_D) \end{bmatrix}$$ (1)
Harris proposed to use the measure as Eq. (2) for interest strength, which combines the two eigenvalues of the auto-correlation matrix in a single measure.

\[ S = \det(M) - \lambda \cdot \text{trace}(M) \]  

No threshold is used in Harris detector, after interest point detection based on interest strength defined as Eq. (2), a non-maxima suppression processing is performed to obtain final interest point result. Harris detector is also a corner detector as FAST detector.

2.3 MSER

MSER (Maximally Stable Extremal Regions) detector is proposed for region detection [10]. A Maximally Stable Extremal Region is defined as a connected component of an appropriately threshold image. The word “extremal” means that all pixels inside the MSER have either higher (bright extremal regions) or lower (dark extremal regions) intensity than all the pixels on its outer boundary. Detection of MSER depends on testing on all possible thresholds and evaluating the stability of the connected components.

2.4 SIFT

SIFT algorithm is proposed by Lowe [1]. It bases on detection local 3D extrema in the scale-space pyramid built with Difference-of-Gaussian (DOG) filters. The DOG representation is obtained by subtracting two successive smoothed images based on continuously smoothing the input image with a Gaussian filter. The local 3D extrema in the pyramid representation determines the localization and the scale of the interest points. For reliability, SIFT eliminates some low contrast key points. Since the DOG gives strong response on edges, an additional filtering step is presented, where the eigenvalues of the full Hessian matrix are computed. The eigenvalues are used to evaluate the localization accuracy of the key points, and the poorly localized key points are removed from the final point list.

2.5 SURF

SURF algorithm was proposed by Bay et al. [2]. It is inspired by SIFT algorithm, box filter based on integral image is used to approximate linear Gaussian filter to speed up the continuously octave generation. Then Hessian operator is employed to determine feature point. After that, a non-maxima suppression processing with a threshold is carried out, and then blob-like feature are obtained.

3 Performance Evaluation Criterion

3.1 Repeatability for Local Feature Detector

The definition of criterion is an important task for performance evaluation. Local feature detector has many properties, while repeatability is the most important for applications such as feature
matching based registration and 3D modelling. Repeatability means a feature detected in one image can be accurately detected in another image. This paper use the repeatability rate defined in [6, 11] as the criterion. Due to the different output of different local feature detector, two kinds of repeatability rate definition are used. One for point feature is defined as [6], and another for affine region is defined as [11]. To calculate repeatability, unique relationship between stereo pair should be a prior knowledge, so this paper uses stereo pair whose relationship can be expressed by homograph transformation as experimental data set, and proposes a method as following section to calculate the homography parameters for generated pyramid images.

3.2 Homography Parameters Relationship Between Pyramid Images

Set $n$ ($n$ is set to 3 in all the experiments) level pyramid image generated from a stereo pair. The scale between succession pyramid images is $r$. Fig. 1 is an illustration of pyramid image with 3 levels, and the scale is used to express the pyramid image level. Such as $r = 1$ means the original image, while $r = 4$ means the highest level of pyramid image. The homography parameters in original images are $(h_1 \ h_2 \ h_3; h_4 \ h_5 \ h_6; h_7 \ h_8 \ 1)$, image coordinate in one of the image is $(x, y)$, while in the downscale image with a scale ratio $r$ is $(x_r, y_r)$, the corresponding coordinate in another image of the stereo pair is $(x', y')$, while in the downscale image is $(x'_r, y'_r)$, so the mapping relationship between original image and downscale image is as Eq. (3):

$$\begin{cases}
    \begin{bmatrix}
        x \\
        y \\
        1
    \end{bmatrix} =
    \begin{bmatrix}
        r & 0 & 0 \\
        0 & r & 0 \\
        0 & 0 & 1
    \end{bmatrix}
    \begin{bmatrix}
        x_r \\
        y_r \\
        1
    \end{bmatrix} \\
    \begin{bmatrix}
        x' \\
        y' \\
        1
    \end{bmatrix} =
    \begin{bmatrix}
        r & 0 & 0 \\
        0 & r & 0 \\
        0 & 0 & 1
    \end{bmatrix}
    \begin{bmatrix}
        x'_r \\
        y'_r \\
        1
    \end{bmatrix}
\end{cases}$$

The relationship of corresponding points between the original pair is expressed as Eq. (4):

$$\begin{bmatrix}
    x \\
    y \\
    1
\end{bmatrix} =
\begin{bmatrix}
    h_1 & h_2 & h_3 \\
    h_4 & h_5 & h_6 \\
    h_7 & h_8 & 1
\end{bmatrix}
\begin{bmatrix}
    x' \\
    y' \\
    1
\end{bmatrix}$$

Fig. 1: Illustration of pyramid image
Substitute the equation (3) to (4), then relationship of corresponding points between the down-scale pair is expressed as Eqs. (5) (6) (7):

\[
\begin{pmatrix}
  r & 0 & 0 \\
  0 & r & 0 \\
  0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
  x_r \\
  y_r \\
  1
\end{pmatrix}
= 
\begin{pmatrix}
  h_1 & h_2 & h_3 \\
  h_4 & h_5 & h_6 \\
  h_7 & h_8 & 1
\end{pmatrix}
\begin{pmatrix}
  r & 0 & 0 \\
  0 & r & 0 \\
  0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
  x'_r \\
  y'_r \\
  1
\end{pmatrix}
\]  
(5)

\[
\begin{pmatrix}
  r & 0 & 0 \\
  0 & r & 0 \\
  1 & 0 & 0
\end{pmatrix}^{-1}
\begin{pmatrix}
  h_1 & h_2 & h_3 \\
  h_4 & h_5 & h_6 \\
  h_7 & h_8 & 1
\end{pmatrix}
\begin{pmatrix}
  r & 0 & 0 \\
  0 & r & 0 \\
  0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
  x'_r \\
  y'_r \\
  1
\end{pmatrix}
\]  
(6)

\[
\begin{pmatrix}
  x_r \\
  y_r \\
  1
\end{pmatrix}
= 
\begin{pmatrix}
  h_1 & h_2 & h_3/r \\
  h_4 & h_5 & h_6/r \\
  rh_7 & rh_8 & 1
\end{pmatrix}
\begin{pmatrix}
  x'_r \\
  y'_r \\
  1
\end{pmatrix}
\]  
(7)

So the homography parameters between a downscaled stereo pair from a stereo pair with known homography parameters can be directly calculated using Eq. (8).

\[
H_r = 
\begin{pmatrix}
  h_1 & h_2 & h_3/r \\
  h_4 & h_5 & h_6/r \\
  rh_7 & rh_8 & 1
\end{pmatrix}
\]  
(8)

4 Experiment Result and Analysis

4.1 Experiment Data

8 image data sets (downloaded from http://www.robots.ox.ac.uk/~vgg/research/affine/) called “bark”, “bike”, “boat”, “graf”, “leuven”, “tree”, “ubc”, “wall” separately as shown in Fig. 2 (a)-Fig. 2 (h) are used for experiments. The image data sets are from stereo image shot a plane scene or the camera only has translation motion. The first image in the left most column of each data set in Fig. 2 is treated as reference image. Homography parameters between the later five images (name 2-6 in following section) and the reference image (name 1 in following section) are given. Then 2 layers pyramid images are generated for each image and the homography parameters between the pyramid images are calculated. The image sizes and characteristic are listed in Table 1.

4.2 Experiment Result

FAST, Harris, MSER, SURF, SIFT algorithm are used to detect interest point and regions on the experiment images and corresponding pyramid images. Then repeatability rates are calculated for these local features, the result is shown in Fig. 3. For each data set, the five detectors are performed, and the results on pyramid images are also plotted in Fig. 3.
Fig. 2: Experiment data set

Table 1: Experiment data description

<table>
<thead>
<tr>
<th>Data set</th>
<th>Image size (r=1)</th>
<th>Image size (r=2)</th>
<th>Image size (r=4)</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>bark</td>
<td>765*512</td>
<td>382*256</td>
<td>191*128</td>
<td>zoom + rotation</td>
</tr>
<tr>
<td>bike</td>
<td>1000*700</td>
<td>500*350</td>
<td>250*175</td>
<td>image blur</td>
</tr>
<tr>
<td>boat</td>
<td>850*680</td>
<td>425*340</td>
<td>212*175</td>
<td>zoom + rotation</td>
</tr>
<tr>
<td>graf</td>
<td>800*640</td>
<td>400*320</td>
<td>200*160</td>
<td>viewpoint change</td>
</tr>
<tr>
<td>leuven</td>
<td>900*600</td>
<td>450*300</td>
<td>225*150</td>
<td>light change</td>
</tr>
<tr>
<td>tree</td>
<td>1000*700</td>
<td>500*350</td>
<td>250*175</td>
<td>image blur</td>
</tr>
<tr>
<td>ubc</td>
<td>800*640</td>
<td>400*320</td>
<td>200*160</td>
<td>JPEG compression</td>
</tr>
<tr>
<td>wall</td>
<td>1000*700</td>
<td>500*350</td>
<td>250*175</td>
<td>viewpoint change</td>
</tr>
</tbody>
</table>

Scale Change. Fig. 3 (a) shows the results for “bark” set as Fig. 2 (a), while Fig. 3 (c) shows the results for “boat” set from Fig. 2 (c). The image transformation is mainly composed of a scale change and rotation. On the higher pyramid level, the repeatability of FAST and Harris detector become higher. SURF detector is also become higher. For SIFT detector on “bark”, the repeatability between 1 and 2, 3 is become higher, while between 1 and 4, 5, 6 become lower, however, they are not low so much. For SIFT detector on “boat”, the repeatability become higher as pyramid image layer become larger. For MSER detector on “bark”, the repeatability between 1 and 3, 4, 5, 6 becomes larger on the second level of pyramid, but in the third pyramid level turns to zero.

Viewpoint Change. Fig. 3 (d) shows the results for “graf” data set as Fig. 2 (d), while Fig. 3 (h) shows the results for “wall” data set from Fig. 2 (h). The stereo image pairs have affine distortion. For the “graf” set, the repeatability of all detectors becomes higher. However, for “wall” set, the repeatability is the same level to some extent, but on the higher level of pyramid image, the repeatability becomes lower. It can be explained that, for texture scene as Fig. 2 (d), the repeatability doesn’t change very much, but for the structure scene as “wall” set, the repeatability turns to lower in the higher level of pyramid image.

Image Blur. Fig. 3 (b) and (f) display the result of “bike” and “tree” data set. As the image is blurred, the repeatability of all the detector become lower, but nearly for all situation, the repeatability on higher layer of pyramid images becomes higher, especially on the highest level of pyramid image.

JPEG Compression. Fig. 3 (g) displays the result of the “ubc” data set. As the compression
rate becomes higher, it nearly has no affection on the FAST and Harris detector. It can be seen that for all situations, the repeatability of higher pyramid image layer becomes higher.

**Light Change.** Fig. 3 (e) displays the result of the “leuven” data set. The light change nearly has no affection on the repeatability of all the detectors on all the pyramid image levels.

![Fig. 3: Repeatability result of the experiment pyramid images](image_url)
5 Conclusion and Remarks

This paper presents a repeatability rate performance evaluation of five common used local feature detector. For texture image, it can be concluded that pyramid image can reduce the affection of scale change and affine distortion to some extent. For structure scene, pyramid image will not change the repeatability too much. The experiment result reveals that it is feasible to downscale large size image to detect local features. Next work will focus on detect evenly distributed local feature from pyramid image.

References