Corner Feature Detection Based on Discrete Spherical Model

Xuebin Qin and Shigang Li
Graduate School of Engineering, Tottori University, 4-101 koyama-mimami, Tottori, 680-8552, Japan
E-mail: qinxuebinhyb@yahoo.com.cn, li@ele.tottori-u.ac.jp

Abstract This paper proposes a method of finding scale-invariant corner feature from full-view image based on discrete spherical model. A full-view image is first mapped to a discrete spherical image. Then, Harris corners are detected in the discrete spherical image. After determining the scale of the found corner points, the descriptor of corner feature is generated from the patch determined by the position and the scale. In comparison with detecting feature from the full-view image directly, the proposed method can find more stable features and achieves higher rate of feature matching for full-view image pairs independent of the feature position at image.

Keywords: full-view image, feature matching, discrete spherical model, descriptor

1. Introduction

Full-view image sensors are widely used in robotics for observing the surrounding scenes. A full-view image sensor usually consists of multiple cameras [1]-[3] to deal with a dynamic environment. Fig. 1(a) shows the full-view image sensor used in our previous study [1], which consists of a pair of fisheye cameras. Both fisheye cameras of the sensor have a max.190 degrees field of view, but point to opposite directions. A sample full-view image captured by the sensor is shown in Fig. 1(b).

Note that in Fig. 1(b) the scene straight lines become curves in the fisheye image, and the curvature of the curves depends on the location of lines in the image. That is, the distortion of features in fisheye image change with their location.

An ideal feature descriptor of full-view image should be invariant to camera rotation. To cope with this problem, we first map the full-view image onto a sphere to obtain a discrete spherical image. The discrete spherical image transformed from Fig. 1(b) in terms of the spherical image format of [1] is shown in Fig. 1(c). In the spherical image, the distortion appearing in the fisheye image is normalized. For example, all scene straight lines are projected as great circles in a sphere. Thus, if we detect and describe features in spherical image, the generated feature descriptor is invariant to camera rotation. A stable feature descriptor will result in a better performance on feature-based motion estimation, scene recognition and so on.

This paper proposes a method of finding scale-invariant corner point from full-view image based on discrete spherical model. A full-view image is first mapped to a discrete spherical image. Then, Harris corners are detected in the discrete spherical image. After determining the scale of the found corner points, the descriptor of corner feature is generated from the patch determined by the position and the scale, at last, features matching is based on the SIFT descriptors. In comparison with detecting feature from the full-view image directly, the proposed method can find more stable features and achieves higher rate of feature matching for...
full-view image pairs independent of the feature position at image.

The rest of this paper is organized as follows. First, in section 2, related research is introduced. The proposed method is presented in section 3. The experimental results are shown and discussed in section 4. The paper is concluded in section 5.

2. Related Research

The detection and description of features is a basic topic in computer vision and pattern recognition. The section below highlights some of the studies that have been carried out in relation to the topic.

2.1 Scale-space image in the form of discrete spherical image

The feature-based matching method is mainly applied for wide baseline images. Generally, the first stage of computation searches over all scales and image locations. The Scale-Space images are obtained efficiently by using convolution of the original image with the Gaussian function. A full-view image is a simple 2D array representation [9]. Hereafter, the image of the spherical image format is abbreviated as the SIF image. The different level subdivisions have different sizes of SIF image. For example, the width and height of the 7-level subdivision SIF image is 2 times the width and height of the 6-level subdivision SIF image, respectively. That is to say, the size of the 7-level subdivision SIF image is 4 times of the size of the 6-level subdivision SIF image. On the other hand, each pixel in the SIF image is a vertex in the geodesic division sphere. Almost all the pixel cells at the vertices are hexagonal in shape. In this research, the advantage of hexagonal is used adequately to obtain scale-space images in the different level subdivision SIF image. The scale-space image \( I(x, y, \sigma) \) at scale \( \sigma \) is obtained by convolution of the SIF image \( I(x, y) \) with the Gaussian function \( G(x, y, \sigma) \):

\[
L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)
\]  

(1)

2.2 Harris corner detector in the SIF image

A new Harris Corner detector is designed in the SIF image. According to the theory of Harris corner detector [10], we need to determine the first-order partial derivatives and integration window. In this research, a hexagonal pixel of a simple averaging differential Sobel operator [14] is used to calculate the first-order partial derivatives in the SIF images. The hexagonal pixel of an average differential Sobel operator is shown in fig. 2. The integration window is implemented efficiently by a hexagonal pixel in the SIF images. However, automatic feature points and corresponding to scale selection may result in redundancy information. The Laplacian operator was used to solve this problem.

2.3 Enhance robust of feature points by Laplacian operator

Automatic scale selection and the properties of the selected scale have been extensively studied in [15]. In order to get robust feature points, we apply Laplacian operator to reduce fake feature points after detecting feature points and scale selection automatically by Harris corner detector. The center of Laplacian operator is symmetry, so it is very suitable for detecting blocked areas. On the other hand, the SIF image is a simple spherical representation. Note that a sphere is isotropic. Therefore an approximation of the Laplacian operator on a hexagonal grid [16] in scale-space of SIF images as Laplacian operator is used. In fig. 3, an approximation of the laplacian operator on a hexagonal grid is obtained by subtracting the value in the center from the average of the six neighboring cells. The result is multiplied by a constant that depends on the grid spacing \( \varepsilon \).

3. Harris Corner Measure and Descriptor

In this section, we first describe how to generate scale-space images in SIF image, and then explain a scale-adapted second moment matrix. Furthermore, in order to get more robust feature points, we use an approximation of the Laplacian operator and at last, SIFT descriptor [6] is used. We explain the proposed method in details as follow.

3.1 The scale-space

In general, the scale-space is obtained by convolution of the original image with the Gaussian function. The property of hexagonal pixel is used. The 12 vertices in the SIF image are ignored because they belong to a pentagon. Each vertex has its hexagonal pixel except the 12 vertices. A hexagonal is considered as a 1-layer Gaussian kernel. A 5-layers Gaussian kernel is formed by extending the hexagonal pixel through

![Fig. 2. Hexagonal pixel of a simple averaging differential Sobel operator. (a) The gradient in the x-direction, (b) The gradient in the y-direction](image)

![Fig. 3. Demonstration of an excellent approximation of the Laplacian operator on a hexagonal grid](image)
their abutment relationship. The outline of the proposed algorithm is given.

Step1. A 5-layer Gaussian kernel is formed by adjacent relationship of hexagonal pixel.

Step2. Calculate the space distance between the center and each adjacent point, and then normalize the space distance as the distance of Gaussian kernel.

Step3. The center pixel value is calculated by convolution of the SIF image with the Gaussian kernel.

For each octave of scale-space, the 6,7,8,9,level subdivision is used. Here, the size of the 6-level subdivision SIF image is 320x128. According to the rule of division, the size of the 7-level subdivision SIF image is 640x256. The others observe the rule.

The scales selected are based on classical SIFT algorithm. First of all, suppose that the circular region of the pair of fisheye sub-image is \( r_f \) pixels. The number of the effective pixels, \( A_f \), included in the two circular regions is computed as follow.

\[
A_f = 2\pi r_f^2
\]

(2)

The size of SIF image is \( A_{si} \), \( i=6,7,8,9 \). The size relationship, \( T \), between SIF image and original image is computed as follows:

\[
T = A_{si} / A_f
\]

(3)

Where SIFT first doubles the image is represented by 9-level subdivision SIF image and pre-smoothes to a start scale \( \sigma = 1.6 \). In order to obtain consecutive scale-space, the set of scale-space images are given in each level subdivision as follows:

\[
\sigma_i = T \sigma \cdot 2^{i/2}, i \in \{-1,0,\ldots,4\}
\]

(4)

3.2 Harris interest point detector in the SIF image

Harris Interest point detector is based on an auto-correlation matrix. The interest points are extracted in an image by measuring the corneriness of a point. The matrix is adapted to scale change and is independent of the image resolution. Here, the derivative as depicted in fig.2 is computed. The integration window is implemented by a hexagonal pixel. The auto-correlation matrix is designed by:

\[
u(X, \sigma_i, \sigma_{si}) = \sigma_{si}^2 G(\sigma_i) \begin{bmatrix} L^2_i(X, \sigma_i) & L_i L_{si}(X, \sigma_{si}) \\ L_i L_{si}(X, \sigma_{si}) & L^2_{si}(X, \sigma_{si}) \end{bmatrix}
\]

(5)

Where \( X = [x, y]^T \) are the coordinates in the scale-space of SIF image. \( \sigma_{si} \) and \( \sigma_i \) are the integration scale and the differentiation, respectively. \( I_i \) and \( I_{si} \) are the SIF image derivatives obtained by the proposed method. The Harris measure combines the trace and the determinant of the matrix. Local maximum of cornerness is the location of interest points. Furthermore, the robust feature points and adapted scale are obtained by Laplacian operator.

3.3 Robust feature points and adapted scale are detected by Laplacian operator

Laplacian operator is very suitable to detect blob area in the scale-space of perspective image[15]. In this study, the Laplacian operator in perspective image is extended to discrete spherical image. On the other hand, Laplacian operator is very sensitive to noise. So Gaussian kernel is used to smooth image and form scale-space image. And then, the SIF image edges are detected by Laplacian operator. That is to say, Gaussian-Laplace operator is used to obtain adaptive block area in scale-space of the SIF image. The Laplacian operator is used to detect adapted area centered at feature point. A series of treatment steps are given as follows.

Step1. Feature points detected by Harris feature detector in scale-space is as initial value.

Step2. Calculate Laplacian response of each point in scale-space by a hexagonal pixel depicted in fig.3.

Step3. Find the local extreme over local scale-space of the Laplacian response for each point depicted in fig.4, otherwise reject the point.

3.4 Feature descriptors

Local descriptors [6]-[8] have been widely researched in respect of image matching. The matching process is implemented by SIFT descriptor in the paper. The descriptor is generated as follows. The feature scale is used to define a circular region centered at extracted interest point. The region generates a perspective image patch. The perspective angle and the image size need to be determined for generating the perspective image. Here, the fixed size 41x41 image patch [4]-[5] was selected. For the perspective angle, the focal length has to be determined and the arc length of covering a spherical model. The feature descriptor is explained in detail as follows.
The effective pixels of a pair of fisheye sub-images are given in (2). Suppose the radius of a sphere is $r_s$ pixels, as in Fig. 5. The number of the pixels of the sphere $A_s$ is:

$$A_s = 4r_s^2$$  \hspace{1cm} (6)

Assuming that the sphere has the same resolution with the pair of fisheye sub-images, the following equation is used to determine the radius of the sphere.

$$r_s = \frac{r_f}{\sqrt{2}}$$  \hspace{1cm} (7)

The arc length is determined by the relationship between the detected scale of feature point and the selected region radius of SIFT descriptor in classical algorithm [6]. An arc length, $I_f$ pixels, is shown in Fig. 5. The perspective angle, $\varphi$, is obtained as the following equation:

$$\varphi = \frac{I_f}{r_s}$$  \hspace{1cm} (8)

![Fig. 5. The detected scale of Harris feature point is used to determine a circular support region on the sphere centered at feature point. According to the relationship between the detected scale and the size of circular region in the classical SIFT descriptor, we are in accordance with SIFT algorithm. The relationship of focal length, perspective angle and perspective image is shown.](image)

A circular region on the sphere is projected to the fixed image plane centered at Harris feature point in Fig. 6. The radius is the adapted scale by our proposed method. The SIFT descriptor is implemented in the right image.

4. Experiments and Results

In the section, we first capture the images by the full-view image sensor shown in Fig. 1(a). The sensor consists of a pair of fisheye cameras which have a max. 190 degree field of view, respectively. A sample of a full-view image is shown in Fig. 1(b). The image size captured by each of the fisheye cameras is 640x480 pixels. The camera parameters including the intrinsic parameters and extrinsic parameters are calibrated beforehand [17]. And then we implement the proposed method in the SIF image shown in Fig. 1(c).

![Fig. 7 Compare to the detected SIF image edge for Laplacian operator](image)

4.1 The effectiveness of Laplace operator in SIF smoothing image

Laplace operator is very sensitive to noise. Generally, the
Gaussian kernel is used to smooth the image, and then use the Laplace operator to detect image edge. The effectiveness of Laplace operator is used in SIF smoothing image. The result is shown in Fig.7. In (a), the SIF image edges are detected by Laplace operator directly. The noise in the image is very disturbing. The spherical projection image of the SIF image is shown in (b). In (c), Gaussian function is smoothed the SIF image (scale=1.5) firstly, and then, Laplace operator is used to detect SIF image edge. The result shows a good performance of noise suppression. The spherical projection image of the image(c) is shown in (d).

4.2 Obtain feature points and adapted scale and then generate SIFT descriptor

For each full-view image, it is first represented by the SIF image. The 6,7,8,9-level subdivision SIF image is used as each octave of scale-space. Feature points are detected by Harris feature detector. The detected feature points and adapted scale regard as initial value. Furthermore, the Laplacian response of each point in scale-space is calculated in SIF image by an approximation of the Laplacian operator on a hexagonal grid. Robust Harris feature and adapted scale are determined by the local extremum over scale-space of the Laplacian response. A sample is shown in Fig. 8.

In the experiment, the radius \( r_s \) of circular region in each sub-image corresponding to the hemispherical field of view is 213 pixels. The radius \( r_p \) of the sphere with about the same resolution is 150 pixels. The arc length, \( l_p \), is obtained by the relationship between the detected scale and the selected region radius of classical SIFT descriptor. The perspective angle, \( \varphi \), is calculated by (8). At last, a 41x41 image patch is generated and get a SIFT descriptor in the image patch.

![Fig.10 Feature matching using the classical method](image)

4.3 Matching result

The two approaches were compared, the proposed method and the classical SIFT method. A pair of the full-view images matching is shown in Fig. 9 by the proposed method in this study. The classical method is used in Fig.10. The number of matching pairs is 120 in Fig. 9 by proposed method, however, 125 pairs are detected in Fig.10 by using SIFT method. The essential matrix E is calculated by the matching pairs and RANSAN to remove outliers [13]. And then, the RANSAC technique is executed in the matching pairs between the proposed method and the SIFT method, respectively. The remaining matching pairs by the proposed method is 97, However, the remaining matching pairs by SIFT method is 61. Outliers in Fig.9 are removed by using RANSAC technique in Fig. 11(a). The remaining matching pairs are considered as the correct matching pairs. The proposed method is verified by 3 arbitrary matching pairs of full-view images in Fig.11(a), (b), (c). The figure shows different scenes. The performance matching between the proposed method and the conventional method are compared in Table 1. For example, in the first row of the table, the number of matching pairs by the proposed method is 120, the number of matching pairs after using RANSAC technique is 97, so the matching rate of the proposed method is 80.8%. However, the matching ratio of the classical method is 48.8%. Finally, the average value of the
number of matching pairs is shown, the number of matching pairs after using RANSAC and correct matching ratio in the table. The correct matching rate of the proposed method is greater than the SIFT method. The experiment shows that the pairs of full-view images have heavy distortion and a big relative rotation. The matching performance by the proposed method is better than the classical method.

4.4 Performance evaluation

To compare the proposed method and the SIFT method, a pair of full-view images in Fig. 9 was employed. The recall and 1-precision were used as evaluation criterion [4]. Recall and 1-precision are defined as follow (9)(10):

$$\text{recall} = \frac{\text{number of correct matches}}{\text{total number of correct matches}} \quad (9)$$

$$1 - \text{precision} = \frac{\text{number of false matches}}{\text{total number of all matches}} \quad (10)$$

To determine if a feature match is correct, epipolar constraints are used under the assumption that the operating environment is rigid. The essential matrix $E$ between image pairs is solved using RANSAC[13] to remove outliers. After finding the essential matrix, the best feature matches are used to reduce the number of initial outliers.

Fig. 12 Recall versus 1-precision result for wide-baseline graphs are shown in (a), (b), (c). The three graphs are consistent with Fig. 11.

The large baseline images are indicated by in Fig. 9. The proposed method and the SIFT method are compared in Fig.
12 using (9)(10). Three graphs are given in Fig.12 and indicate the robustness of the proposed method. In the two formulas, the matching pairs obtained by the proposed method or SIFT method are considered as total number of correct matches. The remaining matching pairs obtained by RANSAC technique between the proposed method or SIFT method are considered as number of correct matches. The essential matrix E is calculated by RANSAC technique in the proposed method. Then for any two features X and X' in each image, a match is considered correct if [X'EX]<threshold. The matching pairs of satisfying the condition are considered as total number of correct matches. Furthermore, in RANSAC technique, the different essential matrix E can be obtained by different thresholds. Figure.12 shows the comparison of matching performance of the two methods. The result shows that the proposed method has better performance for a large relative rotation than the classical method.

5. Conclusions

In this paper, a method of image matching in full-view image is proposed which consists of a pair of fisheye sub-images. In the process, the full-view image is in the form of SIF Image. A Harris feature detector is designed in the scale-space of SIF image. Furthermore, the robust feature points and adapted scale are selected by local extreme of Laplace response in scale-space. These feature points are matched with an adapted SIFT descriptors in SIF image. In the experiment, a pair of full-view image with heavy distortion and large relative rotation is used to show the effectiveness of the proposed method. While the conventional method, which finds SIFT feature directly from the input fisheye image, has the matching rate of 48.8%, our method, which finds SIFT feature from a discrete spherical image, achieved the matching rate of 80.8%. Estimating the full-view camera’s motion from the found SIFT feature is our feature work to do.

References


Xuebin Qin received his BS and MS degrees from Xian University of Science and Technology, China, in 2006 and 2009, respectively. He is now a Ph. D. candidate at Graduate School of Engineering of Tottori University, Japan.
Shigang Li received his BS degree in electrical engineering from Tsinghua University, China, in 1985. He then received his MS and Ph. D degrees from Osaka University, Japan in 1990 and 1993, respectively. After receiving his Ph. D degree, he worked at Osaka University as a Research Associate in 1993. He became an Associate Professor at the Faculty of Information Sciences, Hiroshima City University, Japan in 1995. In 2001, he moved to the Faculty of Engineering at Iwate University, Japan. Since October, 2007, he has been a Professor at the Faculty of Engineering, Tottori University, Japan. His research interests include computer/robot vision, intelligent transportation system and mixed reality.

(Received May 9, 2012; revised September 18, 2012)