Comparative Study of Panoramic Mosaic Technology Based on Fish Eye Image in Urban Monitoring

Mengxue Xu[1, a]

[1]School of General Education, Xi'an Eurasia University, Xi'an 710065, China

[a]xumengxue@eurasia.edu

Abstract

Fish eye lens has short focal length and ultra wide viewing angle range, which can greatly reduce the cost of image acquisition when applied to video surveillance, and has great advantages in panoramic image mosaic. The paper studies panoramic mosaic technology based on fish eye image, compares and analyzes the common algorithms of fish eye image correction and image registration, and selects the suitable algorithm to complete the mosaic of monitoring image.

Keywords

Fisheye image; Panorama Stitching; Image Correction; Image Registration.

1. Overview

Video surveillance has spread to every corner of the city, and the common camera in video surveillance will have monitoring dead angle or blind area, which can not effectively capture a wide range of scenes. Fish eye lens is an ultra wide angle lens based on bionics, with short focal length and ultra wide viewing angle range. It can greatly reduce the cost of image acquisition when applied to video surveillance, and has great advantages in panoramic image mosaic, realizing "all-round" and "real-time" information extraction without blind area[1].

However, in order to obtain super large wide angle, fisheye lens gives up the principle of small hole imaging in the imaging structure, resulting in serious image distortion, so it is necessary to correct the fisheye image in panoramic stitching. The specific implementation steps of panoramic mosaic based on fisheye image are shown in Figure 1. In this process, fisheye image correction, image registration and image fusion are the key modules. This paper focuses on fisheye image correction and image registration.

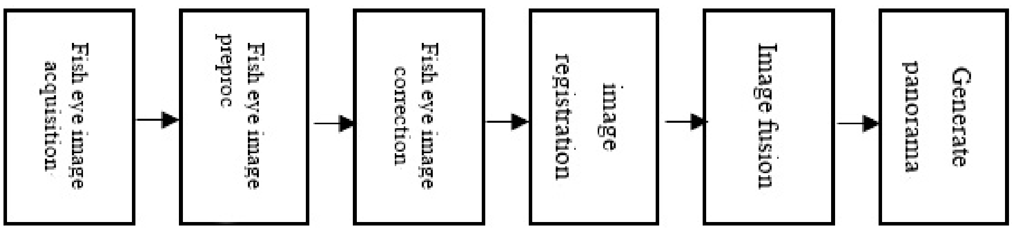

Figure 1. Panoramic mosaic flow chart

2. Fisheye Image Distortion Correction Algorithm

Fisheye lens is a hemispherical lens with convex front. Due to the nonlinear structure, the set distortion of image edge is large, so the edge image of fisheye image is seriously distorted, which is not conducive to direct observation and acquisition of information. [2] Therefore, it is necessary to correct the panorama before mosaic. At present, many scholars at home and abroad have studied the correction methods of fish eye image, and put forward some effective correction algorithms. For example, the correction algorithm based on Scaramuzza, longitude and latitude coordinate method, spherical perspective projection, fast uniform expansion, circle segmentation algorithm, etc. Although the distortion of fish eye image can be divided into barrel shape and pillow shape, the barrel shape distortion is common in reality. In this paper, longitude and latitude coordinate method and spherical perspective projection algorithm are selected for comparative study.

A. Latitude And Longitude Coordinate Method

Longitude and latitude image correction is to restore the abscissa ("long" axis) of the fish eye image pixel to the original position, while the ordinate ("wide" axis) remains unchanged, that is, to map the distorted fish eye image into a normal rectangular image, that is, to reduce the distortion degree of the fish eye image through mapping[3].

The principle of the longitude and latitude image correction method is to use the longitude and latitude lines of the sphere to approximate the deformation of the scene in the fish eye image. Different pixels in each longitude have the same column coordinate value in the corrected image. The greater the longitude, the greater the distortion.[4] Let a (x, y) denote the pixels of the original fisheye image and Al (U, V) denote the pixels of the corrected image:

$$X = X_0 + (U - X_0)\sqrt{R^2 - Y^2}/R\ ,$$
$$Y = V\ .$$

The implementation steps of the longitude latitude coordinate method are as follows:

1) The effective information area of fish eye image is found, the useful image is extracted, the circle center is calculated, and the radius of fish eye image circle is determined;

2) Using the method of longitude latitude mapping, the coordinate system is changed to find out the corresponding relationship between the coordinates;

3) According to the coordinate mapping relationship found in step 2, the pixel information of the source image is assigned to the points on the target image one by one, and the expected correction image is obtained.

Through the analysis of this correction algorithm, it can be found that when the mapping relationship between the distortion point and the corresponding correction point is only on the outermost meridian, the correction effect is the best; on the contrary, when the distortion point is closer to the center of the circle, the fitting effect of the mapping function is worse.

B. Spherical Perspective Projection Method

The spherical perspective projection method uses the fisheye deformation correction model with deformation correction parameters to map the points on the fisheye projection curve of the spatial straight line to the area, and then uses the minimum distance between the spherical point and the big circle to fit the big circle, restore the correction parameters, and realize the correction of the fisheye image， as shown in Figure 2.

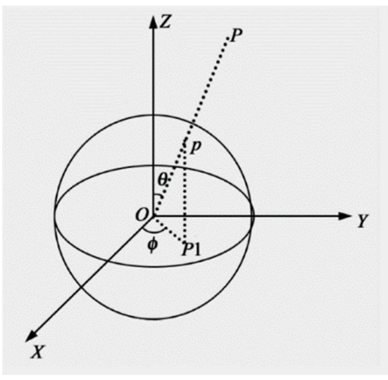

Figure 2. Spherical projection model

Suppose that the coordinates of any point P in space are (x, y, z), connect the origin O and the point P in space to get the ray OP, which is the same as the spherical x 2 + y 2 + Z 2 = R2 intersects point P, and points P are projected onto the XOY plane perpendicular to the optical axis Z of the fisheye lens to obtain the imaging point P1 of the fisheye image. Assuming its coordinates are (U, V), as shown in Fig. 3, the relationship between the coordinates is as follows:

$$\frac{x}{u} = \frac{y}{v} = \frac{z}{\sqrt{R^2 - u^2 - v^2}} = \frac{\sqrt{x^2 + y^2 + z^2}}{R}$$

$$u = \frac{R \times x}{\sqrt{x^2 + y^2 + z^2}}, v = \frac{R \times y}{\sqrt{x^2 + y^2 + z^2}},$$

From the above conclusion, if the pixels of the original fish eye image on the XOY plane are represented by a (x, y), and the corresponding corrected image pixels are represented by Al (U, V), the relationship between them should be as follows:

$$X = X_0 + R \times (U - X_0) / \sqrt{z^2 + (U - X_0)^2 + (V - Y_0)^2}, (z \geqslant R) \ ,$$

$$Y = Y_0 + R \times (V - Y_0) / \sqrt{z^2 + (U - X_0)^2 + (V - Y_0)^2}, (z \geqslant R) \ .$$

The expanded image points and fish eye distortion image points are mapped one by one, and the RGB color value of fish eye distortion image points is mapped to the RGB color value of expanded image. However, the coordinate points corrected by the spherical perspective projection model are not necessarily integer values. Here, the efficient bilinear interpolation algorithm is used to calculate the coordinates of the non integer points.

C. Analysis of experimental results

Longitude and latitude mapping correction algorithm is to convert the nonlinear fish eye image storage into linear longitude and latitude storage, so as to realize image correction; the correction algorithm based on spherical perspective projection first establishes the corresponding mathematical model, and then solves the camera internal, external and correction parameters. From the comparison between Figure 3, Figure 4 and Figure 5, it can be seen that the correction effect of the spherical perspective projection method is more ornamental, but it will lose the edge field of view

3. Image Mosaic Algorithm

After the fish eye image is corrected, it needs to be stitched to form a panoramic image. The essence of fish eye panoramic mosaics is to extract the similar parts of the adjacent image information and quickly find the overlapping area or the best similar overlapping area. The key of image mosaic is image registration technology. In this process, the accuracy of image matching will directly affect the effect of image mosaic.

Figure 3. Original fish eye

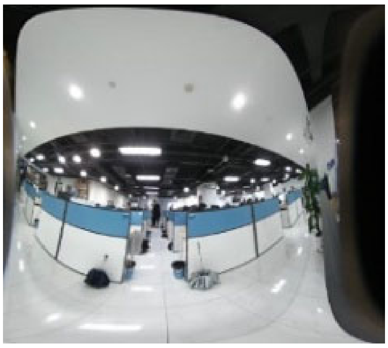

Figure 4. After longitude latitude coordinate method correction

Figure 5. After correction of spherical perspective projection method

Image registration can be divided into two categories: one is to use the region template to traverse to find the region with the highest similarity between images, and then rotate and splice according to the position of similar regions. This method makes full use of all the information contained in the image, and does not need to extract the effective features of the image, but needs to traverse the similarity of each region, so the amount of calculation is large; the other is to use the features in the image（Point, line, edge contour and so on, to obtain high similarity feature matching, to estimate the transformation relationship between images, to carry out image mosaic, which is also a common image registration strategy. The process of feature-based image registration includes image preprocessing, feature extraction, feature matching, transform model estimation and image resampling.

In feature image registration technology, the representative feature extraction algorithms mainly include SIFT, SURF, KAZE and BRISK algorithms. Here, SIFT and SURF algorithms are selected for comparative study.

A. SIFT Algorithm

The core principle of the SIFT algorithm involves detecting key points within the images to be stitched and computing local feature descriptors around these key points, thereby achieving image feature extraction and image matching. There are four steps in feature extraction: Scale space extrema detection, Key point location, Orientation assignment and Feature descriptor construction.

1) Scale space extrema detection: the Gaussian pyramid image is constructed with the help of different scale Gaussian kernel function, and the corresponding score is obtained by subtracting the upper and lower adjacent scale Gaussian pyramid images as the difference of the Gaussian pyramid image [5].

2) Key point localization: The SIFT algorithm employs stability criteria for feature points to filter out a subset of key points, ensuring that the selected key points are stable features.If a point is the maximum or minimum value in the scale space of the Gaussian pyramid image, it can be regarded as the feature point of the image in the corresponding scale.

3) Orientation assignment: After localizing the key points, the SIFT algorithm calculates the gradient magnitude and gradient orientation for all pixels within the neighborhood of each key point.

4) Feature descriptor construction: After determining the location and orientation of the key points, the SIFT algorithm constructs a descriptor in the region surrounding each key point. This descriptor is composed of gradient information from around the key point and exhibits invariance to variations such as illumination and rotation.

SIFT algorithm often chooses 16 × 16 windows in its domain centered on key points. On this basis, it is further divided into 16 4 × 4 sub windows, and each sub window contains vector information in 8 directions, so the final feature descriptor has 128 dimensional vectors.

B. SURF Algorithm

SURF is a classical image local feature detection and description algorithm, proposed by Herbert Bay et al. in 2006. It is actually an improvement of SIFT algorithm. On the basis of retaining the excellent scale, illumination and rotation invariance of SIFT algorithm, surf algorithm improves the extraction and description of feature points. t employs the determinant of the Hessian matrix to locate extrema and utilizes integral images to accelerate convolution operations for generating feature vectors. The Gaussian-filtered Hessian matrix is defined as:

$$H(X,\sigma)=\begin{bmatrix}L_{xx}(X,\sigma) & L_{xy}(X,\sigma)\\ L_{xy}(X,\sigma) & L_{yy}(X,\sigma)\end{bmatrix}$$

By comparing the values of the relevant pixels in the matrix processing with the corresponding 26 feature points in the three-dimensional space, if the point is at the maximum or minimum value, it is regarded as the initial feature point. SURF algorithm is mainly used to ensure rotation invariance with the help of Harr wavelet features in the statistical feature points region. In the feature point area, the sum of the horizontal and vertical Haar wavelet features of all the feature points in the area is counted in 60 ° area, and then the largest sector direction of the wavelet feature is taken as the main direction of the feature point after rotating the area with the same size for one circle [6].

In summary, the SIFT algorithm performs key point detection by constructing a Difference-of-Gaussian (DoG) scale-space and builds high-dimensional descriptors based on gradient orientation histograms. Its design prioritizes extremely high feature distinctiveness and robustness as core advantages, maintaining excellent matching accuracy under various complex imaging conditions. As a result, it has become a classic benchmark in the field of feature extraction. However, its relatively high computational complexity limits its application in scenarios with stringent real-time requirements.

As an accelerated improvement over SIFT, the SURF algorithm achieves a significant enhancement in computational efficiency while retaining sufficient robustness. Its key technical contribution lies in utilizing integral images and box filter approximations to rapidly compute the Hessian matrix for key point detection and generate descriptors based on Haar wavelet responses. This series of engineering optimizations enables SURF to be typically an order of magnitude faster than SIFT, making it particularly suitable for real-time or near-real-time vision systems with strict computational time constraints.

C. Experimental Results and Analysis

This study conducted a quantitative comparison of the core performance of the SIFT and SURF algorithms in surveillance image registration tasks. As shown in Table 1, a systematic evaluation was performed from three dimensions: the number of detected feature points, matching accuracy, and computational efficiency. The results clearly reveal the fundamental differences between the two algorithms in terms of design philosophy and performance characteristics. A detailed analysis is as follows: On the same test image, the number of feature points detected by the SURF algorithm (406) was significantly lower than that detected by SIFT (524). This phenomenon is directly related to their core detection mechanisms. SIFT constructs a fine-grained Difference-of-Gaussian (DoG) scale-space, enabling it to capture denser, multi-scale local extrema. In contrast, SURF employs a fast detector based on an approximate Hessian matrix. While enhancing speed, it tends to filter for feature points with higher contrast and greater stability, resulting in a relatively smaller yet more concentrated set of points.

Regarding matching performance, the SIFT algorithm demonstrated a matching accuracy of 87.6%, notably higher than SURF's 81.3%. This verifies that the 128-dimensional gradient orientation histogram descriptor used by SIFT possesses stronger distinctiveness and discriminative power, maintaining higher robustness against complex scenes and detailed variations. However, this high precision comes at the cost of significant computational overhead. As indicated in Table 1, the matching process for SIFT took 112.4 seconds. When processing image pairs with large overlapping areas, the vast number of feature points it generates drastically increases the complexity of the description and matching stages, severely impacting the algorithm's timeliness and making it difficult to meet real-time processing requirements.

Table 1. Experimental data of SIFT and SURF Algorithm

Algorithm	Number of feature points detected	Matching accuracy	Matching time (/s)
SIFT	524	87.6%	112.4
SURF	406	81.3%	37.6

In contrast, SURF excels in efficiency. Its matching time was only 37.6 seconds, approximately 3 times faster than SIFT. This advantage primarily stems from two key optimizations: first, the use of integral images to achieve constant-time computation for box filters, and second, the adoption of a compact descriptor (typically 64-dimensional) based on Haar wavelet responses. These design choices allow it to avoid the time-consuming iterative scale-space construction and fine-grained gradient computations inherent in SIFT, achieving a leap in processing speed. Although the discriminative ability of its descriptor is slightly weaker, leading to a modest decrease in matching accuracy, its performance remains within an acceptable range for the frame-to-frame registration common in surveillance scenarios.

In summary, the experimental data quantitatively corroborate the theoretical analysis: The SIFT algorithm demonstrates excellent performance in terms of scale, rotation, and affine invariance, offering high matching accuracy. However, it is computationally complex and time-intensive, making it suitable for non-real-time scenarios where precision is paramount. Through clever engineering optimizations, the SURF algorithm achieves an order-of-magnitude improvement in computational efficiency at an acceptable cost in accuracy. Therefore, for urban surveillance image stitching applications with stringent real-time requirements, SURF holds a distinct practical advantage due to its faster processing speed. Ultimately, the choice between algorithms represents a trade-off between "accuracy" and "efficiency," dictated by the specific requirements of the application.

4. Conclusion

Panoramic mosaicing technology based on fish eye image has a great application prospect in urban monitoring. In this paper, the key modules of panoramic mosaicing process are fish eye image correction and image registration. The corresponding algorithms are selected for comparative study. Through experimental analysis and comparison of the characteristics and advantages and disadvantages of each method, the spherical perspective projection method is selected to correct the fish eye image, and surf algorithm is used to correct the fish eye image Finally, the panoramic mosaic is completed. Of course, the above research has some limitations and needs to be further improved. With the development and application of deep learning and graph neural network, it will greatly promote the image mosaic technology, and the operation speed and accuracy of image mosaic will be greatly improved.

References

[1] He, Z., & Zhang, J. (2020). A method for distortion correction of fish-eye images based on stitching algorithm. Modern Computer, (03), 57.

[2] Cui, H., Chen, J., & Dayu. (2008). Research and implementation of fisheye image correction and splicing. Computer Engineering, 33(10), 190–192.

[3] Li, Y. (2024). Research and implementation of panoramic surround view method based on six cameras [Doctoral dissertation]. Jilin University.

[4] Yao, B., & Bai, C. (2020). Survey of feature based image registration technology. Graphic Image, (20), 52–54.

[5] Zhen, Q., & Huang, D. (2020). Performance comparison of SIFT and SURF in UAV image matching. Information Recording Materials, 21(8), 220–221.

[6] Xiao, L., & Liu, W. (2018). Hyperspectral polarization image registration method based on image distortion correction. Infrared Technology, 40(8), 776–778.