Research on Panoramic Image Registration Approach based on Spherical Model

Xiaohui Wang[1], Kehe Wu[2] and Shengzhuang Wang[3]

[1]*Postdoctoral Mobile Research Station of Management Science and Engineering, North China Electric Power University, Beijing 102206 P. R. China*
[2] *School of Control and Computer Engineering, North China Electric Power University, Beijing 102206 P. R. China*
[3] *School of Control and Computer Engineering, North China Electric Power University, Beijing 102206 P. R. China*

[1]*wxh258@126.com,* [2]*epuwkh@126.com,* [3]*wangshengzhuang@126.com*

Abstract

Panoramic image registration is the core content of virtual three-dimensional system based on images. According to the shortage of fisheye images registration using the traditional image registration method, this paper put forward strategy of panoramic image registration, which was divided into three steps: project the fisheye images on the sphere, extract and match the feature points, transform the coordinates of panoramic images. The focus was improving the SIFT algorithm and RANSAC algorithm for image feature exaction and matching, whose target was to meet the demand of the fish-eye image registration. Finally, experiments were designed to verify the effectiveness of supposed algorithm and fisheye image registration, and the future direction of panoramic images registration was also pointed out.

Keywords: Spherical Model, Image Registration, Feature Point, SIFT, RANSAC, Coordinate Transformation

1. Introduction

With the development of virtual reality technology, three-dimensional panoramic technology based on images has been widely used, and panorama generation technology has also become research focus of digital image processing. Panoramic images, which were generated as collecting fisheye images, registering panoramic images and three-dimension rendering, could be used to build the virtual scene and ensure the realism and integrity of the scene. In the whole process, panoramic images registration is the most important, which not only relates the generating efficiency of panorama, and also affects the scene rendering.

Traditional image registration was completed in the planar model, first find the matching features or regions between pending images, and then match the features in the planar for image fusion. Such image registration technology contains: method based on region [1], method based on phase [2], method based on feature [3]. Professor Szeliski from Microsoft Research supposed an eight-parameter model for projection transformation [4], after that experts optimized the 8 parameters to 4 parameters with L-M algorithm. Mr. Brown reported an article named Recognizing Panoramas in 2003 ICCV Conference, which supposed an image registration algorithm based on SIFT. The algorithm had invariance in the affine, scale

and illumination, and combined the low frequency and high frequency information in the image effectively, which reduced the loss of image details.

Fisheye images have the characteristics of wide viewing angle and seriously anamorphic edge, so the traditional image registration method needed to be improved. The process contains the following steps:

(1) Make the fisheye image to three-dimensional that is projecting it to the sphere.

(2) Extract and match the feature points in the 3D scene.

(3) Generate the panoramic image with coordinate transformation.

2. Spherical Projection of Fisheye Image

According to imaging characteristics and feature of the fisheye images, the fisheye images needed to be mapped before image registration, which could be divided into cube mapping, cylindrical mapping and spherical mapping due to the different projections. Spherical mapping is the inverse process of the fisheye imaging; it could express the panoramic space completely, and reflect any direction in the space of the scene. Though the intersecting and positioning algorithm is more complicated, the projection reflects the visual characteristics of the human eye, and describes viewpoint space in the most natural form, so this paper used spherical mapping to project the fisheye images.

The process of spherical mapping for the fisheye images is to project the multiple pending mosaic images onto a spherical model, and to store it with three-dimensional spherical coordinates. This form not only eliminates the redundant information that may exist between the original images, but also acquires azimuth information of the each pixel on the image in the view space.

The imaging and spherical mapping of fisheye images relate to three coordinate systems, which contain world coordinate system XYZ, camera coordinate system xyz and plane coordinate system UV, as shown in Figure 1. Set the camera direction as (α, β), arbitrary point in the real world as P(x, y, z), imaging point in the fisheye image as P'' (u, v), spherical projection points as P'(x', y' z'). According to the imaging characteristics of fisheye lens, we can get the mapping relationship as formula (1).
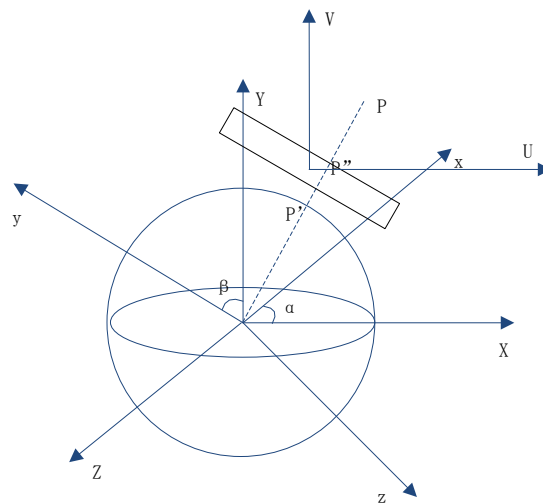

Figure 1. Fisheye image forming and spherical projection

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\alpha & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & -\cos\alpha \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad (1)$$

Mapping of the XYZ coordinate system and the UV coordinate system: take (0,-R, 0) as a starting point and (0, R, 0) as the end point, the Semicircle along the direction of the sphere for the V-axis, the U-axis where the former point (x, y, z) on the V-axis take the point (R, y, 0) as the starting point and the end point, which is a counter clockwise direction circle. So the corresponding relationship between P' and P" is as follows, the y coordinate of P' is as formula (2), the x coordinate of P' is as formula (4), the z coordinate of P' is as formula (5), r is defined as U axis radius of P', which is shown as formula (3).

$$y' = -\cos(v \times \pi) \times R \qquad (2)$$

$$r = \sqrt{R^2 - y'^2} \qquad (3)$$

$$x' = r \times \cos(2\pi \times u) \qquad (4)$$

$$z' = r \times \sin(2\pi \times u) \qquad (5)$$

From above formula, we can obtain the relation of p' and p", which is as formula (6).

$$\begin{cases} u = R \cdot arc\,\cos\left(\dfrac{x\cos\alpha + y\sin\alpha\sin\beta + R\sin\alpha\sin\beta}{\sqrt{x^2 + (y\sin\beta + R\cos\beta)^2}} \right) \\ v = R \cdot \left[\dfrac{\pi}{2} + arc\,\tan\left(\dfrac{y\cos\beta - y\sin\beta}{\sqrt{x^2 + (y\sin\beta + R\cos\beta)^2}} \right) \right] \end{cases} \qquad (6)$$

At this point, the relationship among the world coordinate system, camera coordinate system and the plane coordinate system were established, and the point of reality scene, fisheye image and projection sphere could be one-to-one transformation.

3. Image Feature Extraction and Matching

Image registration based on feature contains three steps:

(1) Calculate and extracts the feature set of the image.

(2) Calculate the positional relationship between the mosaic images with the feature set.

(3) Transform the mosaic image with the positional relationship, and convert to the unified coordinate system.

At present, common characteristic information includes characteristic contour, characteristic curves; characteristic points (corner points, high curvature points, largest changes in the gray scale value point, *etc.*). Matching method with feature point is one of the most popular method, Harris [7], KLT [8], SIFT [9] are the common algorithm. SIFT, which has better image scaling, rotation invariance, fit for the fisheye images, we mend the SIFT algorithm, improve the feature point extraction efficiency, and make it more suitable for fisheye images registration requirements.

3.1. Feature point extraction based on SIFT

Feature point extraction of SIFT involves four steps: (1) detect extreme of scale space; (2) accurate position of feature points; (3) determine the main direction of feature points; (4) generate SIFT feature descriptor.

(1) Detect extreme of scale space

Search all positions of the images on different scales space and identify the potential feature points with scale and rotation invariant by Gaussian differential function, whose main idea is to introduce a scale parameter in image information processing model, and obtain the scale space representation sequence by the continuous change scale parameters, then extract the main contours of these sequences, which is used as feature vectors for edge, corner point detection and feature exaction of different resolution. Scale space of images L(x, y, σ) is defined as the convolution of scale-changed Gaussian function G (x, y, σ) and the original image I(x, y), shown as formula (7).

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{7}$$

Wherein, * represents the convolution of the x and y directions, (x, y) represents an image pixel position, σ is the factor of scale space, the smaller value makes the smoother image and the smaller scale. The large scale corresponds to the profile characteristics of the image, while the small scale corresponds to the detail of the image characteristics. G (x, y, σ) is the Gaussian differential function, defined as formula (8).

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{8}$$

Scale space use Gaussian pyramid in achieving, the pyramid construction is divided into two parts:

① Down-sample the images and get the image under different scales.

② Carry out Gaussian blur on the image under different scales.

Pyramid model of image refers as a tower model, which is formed down-sample constantly from the original image, the original image is the first layer of the pyramid, and the layer adds as every down-sampling. The level of pyramid is determined according to the original size of the image and the size of the top image, formula (9) is the process.

$$n = \log_2\{\min(M,N)\} - t, t \in [0, \log_2\{\min(M,N)\}] \tag{9}$$

Wherein, M, N is the size of original image, t is the log of top level image.

In order to detect the extreme points of the scale space, each pixel must be compared with adjacent pixels, which not only contain the points in the same scale, but also the vertically adjacent pixels, and then determine whether the point is the extreme point of its neighborhood, if its gray value is the maximum value and then recorded as candidate extreme points.

(2) Accurate position of feature points

In order to obtain accurate feature points, we fit candidate extreme points by 3-dimensional quadratic function, simultaneous removed the low contrast feature point and the unstable

candidate point on the edge, finally, match the feature points of stability and anti-noise capacity.

① Remove low contrast feature point

DOG function is expanded with Taylor expansion in the scale space (fitting function) as formula (10), $X = (x, y, \sigma)^T$, derivative of the equation and make it equal to 0, we can get the offset of the extreme points as formula (11), and extreme point equation is like (12).

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \qquad (10)$$

$$\hat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X} \qquad (11)$$

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} \qquad (12)$$

Wherein, $(x, y, \sigma)^T$ represents the excursion offset interpolation center, when the excursion is greater than 0.5 in any dimension, it need to change the location of the point of the current key because the interpolated center has been shifted to its neighboring points, while to interpolate at the new location again and again until convergence. If the number of iterations exceeded the set value or the point was beyond the scope of the image boundary, the point should be deleted.

② Remove the unstable candidate point on the edge

Due to the DoG operator will produce a strong response to the edge, so it need to remove the unstable critical point on the edge in order to get a stable critical point, the principle of which is that the edge feature points are larger in the peak of the curvature of the Gaussian function and at the same time the vertical direction of the curvature value is smaller. The main curvature can be determined by a 2x2 Hessian matrix H, as formula (13), characteristic of H (α and β) represents the the gradient of x and y direction, Tr(H) represents the trace of matrix H, which can be obtained by the sum of the diagonal elements of matrix H, Det(H) represents the determinant of the matrix H, Tr(H) and Det(H) is calculated by formula (14) and (15).

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \qquad (13)$$

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \qquad (14)$$

$$Det(H) = D_{xx} D_{yy} - (D_{yy})^2 = \alpha\beta \qquad (15)$$

Assume α is the larger characteristic value, and β is smaller, $\alpha = \gamma\beta$, formula (16) shows that the main curvature of D the characteristic values of H is proportional, the ratio is smallest when the two characteristic values are equal, and becomes greater as γ increases, which means it is a edge point. Setting a threshold value r, if the value of the formula (16) is less than r, the key points will be retained, otherwise, name it the corresponding point of edge, do the culling process.

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(\gamma\beta + \beta)^2}{\gamma\beta^2} = \frac{(\gamma + 1)^2}{\gamma} \qquad (16)$$

(3) Determine the main direction of feature points

In order to make the descriptor rotation invariance, it need to confer a main direction for every feature point using the statistical information of the neighborhood pixel point. The main direction can be collected as gradient and direction distribution characteristics in the Gaussian pyramid image 3σ neighborhood pixel for the key points detected in the DOG pyramid. The gradient and direction at point(x, y) is shown as formula (17) and (18).

$$m(x, y) = \sqrt{\left(L(x + 1, y) - L(x - 1, y)\right)^2 + \left(L(x, y + 1) - L(x, y - 1)\right)^2} \qquad (17)$$

$$\theta(x, y) = \tan^{-1}\left(\left((L(x, y + 1) - L(x, y - 1))\right)/(L(x + 1, y) - L(x - 1, y))\right) \qquad (18)$$

The method of calculating the main direction of feature point is as follows. Sample a certain neighborhood window centered by feature point, gather the gradient direction of the pixel in the window, and express with gradient histogram, which ranges of 0-360 degrees(10 degrees intervals to determine a direction), the peak represents the main direction of the feature point in the gradient histogram. In order to enhance the robustness of the matching, a secondary direction is set, which is the direction 80% greater than the main direction, therefore, feature point may have a plurality of directions.

(4) Generate SIFT feature descriptor

Through the above steps, the feature point has information of scale, position, and the main direction. In order to improve the correct rate of subsequent feature point matching, it need to establish a descriptor for each feature point, i.e. use a set of vectors to represent the feature point descriptor, which can ensure that it has a high stability, and does not vary with illumination, view angle or other changes.

Feature point descriptor is a unique vector generated by statistics on the neighborhood of the pixel gradient information of feature points, whose method is as follows: (1) rotate the axis to the main direction of a stable feature point, to ensure it rotation invariance, (2) extract 4*4 window centered by feature point, wherein each sub-window contains 4*4 pixels, (3) calculate the gradient histogram of eight directions for each sub-window, sort by scale of feature point, and finally form a 128-dimensional feature vector(4*4*8), which is the feature point descriptor.

3.2. Feature point matching based on RANSAC

Feature points in different images need to be matched to determine whether the feature point is similar after the feature point's descriptor were generated. Because feature points descriptor contains the entire characteristic, the neighborhood information can be define the similar between feature points, and Euclidean distance is the common method. Assume feature points descriptor of p and q respectively show as Des_p and Des_q, the Euclidean distance can be defined as formula (19).

$$d = \sqrt{\Sigma_{i=0\dots127}\left(Des_p(i) - Des_q(i)\right)^2} \qquad (19)$$

Wherein, i represents the dimension of feature point descriptor. The integrated Euclidean distance d can be obtained as adding the Euclidean distance of each dimension, and the similar criterion is by comparing with the set threshold value.

Search from the given feature points with k-d tree[10], we can obtain the nearest feature point P_1 and the second nearest feature point P_2 of feature point P. Calculate the ratio of Euclidean distance between P, P_1 and P, P_2, and compare it with the set threshold value to determine whether it successful or not.

After the feature points were matched with SIFT algorithm, there still existed incorrect matching points, and it needed to adopt geometric constraints or additional constraints to eliminate the error matching points. RANSAC algorithm [11] is a good method. RANSAC algorithm is used to get the correct sample data by calculating mathematical model parameters based on a set of sample data sets contain abnormal data, it could retain effective match points, and eliminate error matching points. The algorithm is currently used in the error data processing for a great deal of areas with a computing speed, good stability advantages.

RANSAC algorithm divides the given data into effective data and the error data, and corrects the two parts of data set by calculating until getting the final data set. Take two-dimensional linear data set for example, at first the algorithm search two random points in the data set, form a straight line by connecting the two points, and judge the distance between other points and the line to divide valid points and invalid point; secondly, concentrated in the valid data, and to re-fit the data using the least squares method to produce a straight line, to re-set threshold value, determine whether the new valid data sets; repeated iterations until point data set does not change, then the algorithm is ended. The algorithm itself is an iterative process, to look for valid data point and exclude unusual data points. This paper designs the following process with RANSAC algorithm for the matching feature points.

① Select four random pairs of feature points, and calculate to get transformation matrix M of 3*3 as initial model;

② Assign a distance threshold value D, and take all the matching points into the matrix M and calculate the distance. Mark it interior point if the distance is within the threshold, otherwise outer point. If the data sets is stable in a non-initiating calculation proceeds, skip to ④, else to ③;

③ Re-calculate the matrix M for all matching pairs;

④ End the calculation and return all valid data sets.

4. Coordinate Transformation of Panoramic Image

The transformation matrix of the image coordinates has different forms in different parameter model; the projection model is the same, while the rotation transformation parameters are different. Assume I_1, I_2 represents two images, f_1, f_2 represents its focal length, the center of the image represents the projection center, marked as (c_x, c_y), M represents the coordinate transformation matrix from I_1 to I_2.

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = K_1 R_1 \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad (20)$$

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = K_2 R_2 \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad (21)$$

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \cong K_2 R_2 R_1^{-1} K_1^{-1} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = M \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \qquad (22)$$

$$K_i = \begin{bmatrix} f_i & 0 & c_x \\ 0 & f_i & c_y \\ 0 & 0 & 1 \end{bmatrix} \qquad (23)$$

Assume the camera parameters of I_1 and I_2 are$(\alpha_1, \beta_1, \gamma_1),(\alpha_2, \beta_2, \gamma_2)$, matrix M represents as direction angle, so matrix M can be represents as formula(24), and the coordinate transformation is as formula (25).

$$M = K_2 R_2 R_1^{-1} K_1^{-1} \qquad (24)$$

$$R_i = \begin{bmatrix} \cos\gamma_i & \sin\gamma_i & 0 \\ -\sin\gamma_i & \cos\gamma_i & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta_i & 0 & \sin\beta_i \\ \sin\alpha_i\sin\beta_i & \cos\alpha_i & -\sin\alpha_i\cos\beta_i \\ -\cos\alpha_i\sin\beta_i & \sin\alpha_i & \cos\alpha_i\cos\beta_i \end{bmatrix} \qquad (25)$$

5. Experiment Results

In order to validate algorithms of image registration, this paper designed experimental environment based on actual application scenario. Figure 2 shows four fish eye images with the same viewpoint and different directions in the substation.
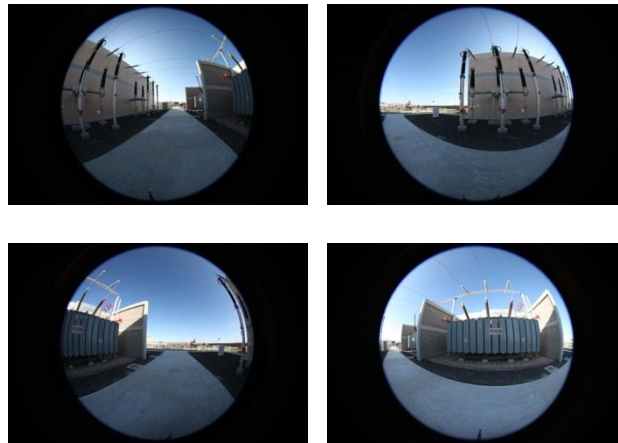
Figure 2. Original fisheye images

Step 1 Project the original fisheye image onto sphere, that is to say, each fisheye image projected onto the inner surface of the hemisphere. In order to show the projection result, we mapped result to plane based on coordinate transformation of panoramic image, as Figure 3.
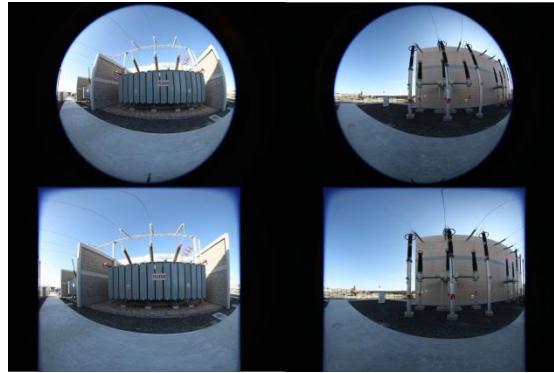
Figure 3. Plane effects of fisheye images

Step 2 According to the projected images, extract the feature points based on SIFT algorithm, shown as left part of Figure 4, and calculate the feature point descriptor for all the extracted feature points, shown as right part of Figure 4.

Figure 4. Feature points and its descriptor

Step 3 Match the feature points for the adjacent fisheye images based on the feature point descriptor and its form is connecting the matched feature points, shown as Figure 5.
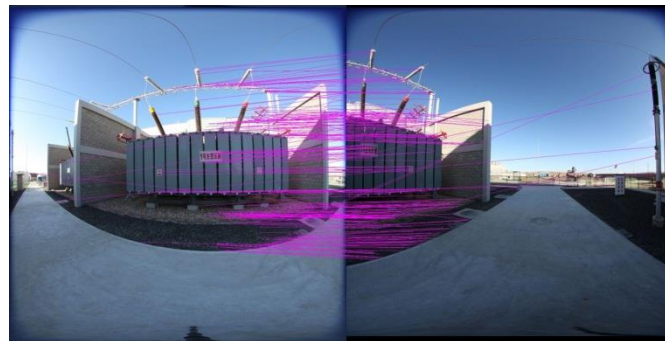
Figure 5. Matching results of feature points

Step 4 Mosaic the four fisheye images of the same viewpoint and form the panoramic image, which just fills the inner surface of the sphere. In order to show the results, the panoramic image was mapped onto the plane, shown as Figure 6.

Figure 6. Panoramic image

The experimental results showed that Spherical projection of fisheye image and Coordinate transformation of panoramic image could effectively achieve the spherical image and the image plane mapping, at the same time, feature point extraction based on SIFT and feature point matching based on RANSAC could quickly find, match and mosaic the images. The panoramic image generated based on these algorithms is clear, non-distortion, and containing information of the whole scene. Therefore, the proposed algorithms are real and effective, and they can provide theoretical support for the virtual three-dimensional system of substation.

6. Conclusion

Panoramic image registration method based on a spherical model is the core content of the virtual three-dimensional system based on images. The paper designed the spherical mapping model of the image, extraction and matching algorithm of fisheye image feature point based on SIFT and RANSAC algorithm, and also panoramic image coordinate conversion algorithm. However, there are still requirements to improve the research, the following shows the future research directions.

① Efficiency of the fish-eye image correction, the current correction algorithm is due to calculate the image critical points and finish the secondary correction, so that when the image is larger, the correction time is longer.

② Determining whether the two images matching, this paper decided it according to matching pairs and total pairs of feature point, what could cause mis-match problem, so a better model is needed.

③ Memory problems of high resolution images, Memory leaks could lead to mismatch, so a memory optimization model needed to be designed.

References

[1] H. Zhang and D. Cui, "Study and Implementation 0f Algorithm in Creating Panoramic Image", Computer Engineering, vol. 28, no. 6, **(2013)**, pp. 95-96.
[2] X. M. Li, X. P. Zhao, L. Zheng and Z. Y. Hu, "An Image Registration Technique Based on Fourier-MellinTransform and Its Extended Applications", Chinese Journal of Computers, vol. 29, no. 3, **(2006)**, pp. 466-472.
[3] M. F. Cohen, D. Salesin and R. Szeliski, "Photographing long scenes with mufti-viewpoint panoramas", ACM Transaction on Graphics, vol. 25, no. 3, **(2006)**, pp. 853-861.
[4] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, **(2003)**.

[5]   M. Brown and D. G. Lowe, "Recognizing Panoramas", Proceedings of the 9th International Conference on Computer Vision, **(2003)**.
[6]   X. Wang, K. Wu and Y. Cheng, "Research on Virtual 3D Station based on Images", Appl.Math.Inf.Sci., vol. 7, no. 1L, **(2013)**, pp. 225-231.
[7]   C. J. Harris and M. Stephens, "A combined corner and edge detector", Proceedings of the 4th Alley Vision Conference, **(1988)**.
[8]   J. Shi and C. Tomas, "Good Feature to Track", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **(1994)**.
[9]   D. G. Lowe, "Distinctive image features from scale-invariant key points", International Journal of Computer Vision, vol. 60, no. 2, **(2004)**, pp. 91-110.
[10] S. Arya and D. M. Mount, "An optimal algorithm for approximate nearest neighbor searching", Journal of the ACM, vol. 45, no. 6, **(1998)**, pp. 891-923.
[11] Q. Tianwei, A. Bo and C. Guilan, "Application of improved RANSAC algorithm to image registration", Journal of Computer Applications, vol. 07, **(2010)**, pp. 1849-1851.

Authors
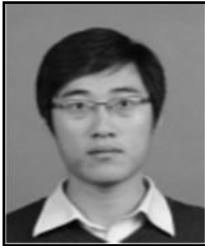
Xiaohui Wang. He received his PhD in Electrical Engineering (2012) from North China Electric Power University. Now he is a Postdoctoral of Management Science and Engineering of North China Electric Power University. Since 2011 he is Member of CSEE. His current research interests include different aspects of GIS and Electric power information technology.

Kehe Wu. He received his M.E. in Computer (1995) and PhD in Thermal Engineering (2009) from North China Electric Power University. Now he is full professor at Control and Computer Engineering School, North China Electric Power University. Since 2009 he is executive vice president of the university. Since 2007 he is member of CAAI. His current research interests include different aspects of Artificial Intelligence and Electric power information technology.

Shengzhuang Wang. He graduated from Computer Science and Technology (2011) from North China Electric Power University. Now he is a postgraduate of Control and Computer Engineering of North China Electric Power University. His current research interests include different aspects of GIS and Workflow.