International Journal of Multimedia and Ubiquitous Engineering
Vol. 10, No. 1 (2015), pp. 47-58
http://dx.doi.org/10.14257/ijmue.2015.10.1.5

Key Point Detection in 3D Reconstruction Based On Human-Computer Interaction

Zhu Shi Wei[1], Zhang Xiao Guo[2], Lv Jia Dong[3] and Wang Qing[4]
[1,2,3]*Southeast University Southeast UniversitySoutheast University*
[4]*Southeast University*
zhu_shi_wei@126.com iszhang@yahoo.comxyc_2028@163.com
W3398a@263.net

Abstract

Aiming at solving problem of points' redundancy caused by full automatically detecting points and the problem of large workload caused by picking all points manually, I advanced a new method of picking points which is based on Human-Computer Interaction in our 3D reconstruction platform after automatically detecting points. We first detected and matched points automatically and got the homograph matrix between two images, then projected points which were picked by hand on the one image to the other image, at last we would search the interesting feature points in the neighborhood of corresponding points in the two images. Experiments have shown that this method decreases the redundancy brought by large number of points and successfully finds the important feature points, so it lays a good foundation for 3D reconstruction.

Keywords: *3D reconstruction; Human-Computer Interaction; Feature Points*

1. Introduction

IBM (image based modeling) [1] is a techniquerecovering three dimension coordinates of contour points of real scene from images to get realistic model with computer vision principle. However, not all the contour points are necessary, only those which are distinctive from their neighborhood points count .We call these points feature points.

The method of extracting feature points is of blind, it will extract all the points in the images which meet the judgment of feature points. But not all the feature points will make sense in 3D reconstruction. For example, many disorderly and confused points around the target object will be detected. These points are not we need, and they increase the amount of computation in 3D reconstruction. In addition, a large number of feature points on the target object will be detected which are redundant. We do not need so many points in reconstruction. They will also increase the amount of computation.

In order to solve these problems, we can modify the judgment of the feature points and raising the threshold of judgment to reduce the redundancy of feature points. However, it may result in omitting some important points which can reflect the geometrical feature of target object. Thus, it will affect the effect of reconstruction ultimately.

So we introduce a method based on human-computer interaction which takes into account of automation of detecting and matching, time complexity, space complexity of reconstruction. Combining with the point searching in small region, it solves the problem of large amount of computation brought by the point redundancy. At the same time, it can also avoid situation that some important feature points in the images fail to be automatically discovered.

2. The Whole Progress of Algorithm

The Algorithm mainly includes three modules: detecting and matching of image feature points, automatically obtaining the transformation matrix between images, points searching in small region.

In the first module,feature points detecting and matching directly affects the subsequent image transformation matrix and the points searching in the small area. As a result, choosing what kind of feature extractors and how to remove the wrong points matching isvery important. Besides, it should be noticed that setting the appropriate threshold value to control the number of feature points to avoid lots of meaningless feature points be extracted is important.

In the second module, we can get the transformation matrix between the images according to the matching points which are got from the first module. However, as the matching points which have removed some mismatching points are still not entirely correct, and not all feature points can meet a transformation matrix because of the different depth of points, we cannot easily get relatively right matrix. Thus we use the RANSAC (random sampling consensus) to get the matrix which can meet the most of the matching points.

In the third module, as some important feature points are not discovered automatically because of the high threshold set in the first module, it affects the performance of 3D reconstruction directly. So we should search the key points in small area based on human-computer interaction to obtain the points we want.

Flow chart of the whole algorithm is shown below:

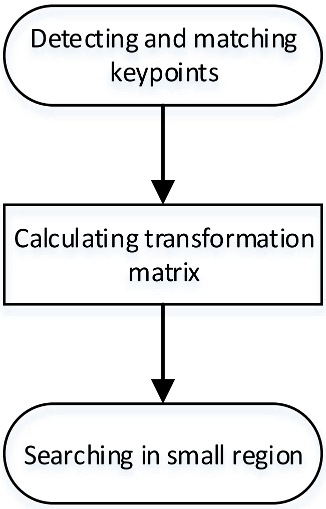

Detecting and matching keypoints
Calculating transformation matrix
Searching in small region

Figure 1. Algorithm Flow Chart

3. Feature Point Detecting and Matching Algorithm

Image feature point extraction and matching is the key step in the 3 D reconstruction. In practical problems, the image could be affected by noise, background interference and could be changed by light, scaling, rotation, affine. Therefore, we should choose a reasonable image feature point detector to make these feature points not only noise resistant but also unchanged under the above changes. Current methods of the local invariant feature points detector are mainly SIFT [2] (Scale Invariant Feature Transform) and SURF [3] (Speed up Robust Features). SIFT detector makes full use of the properties of scale space to get extreme value point in the scale domain and space domain as the feature point. At same time, it uses the feature point’s scale to determine the feature area which can be used to describe the feature point. This method greatly solve the problem of location and size of feature area. SURF detector which is an improved method of SIFT was advanced by Herbert Bay in 2006. The study found that SURF is rotation invariant and resistant to image blur, at the same time, its computing speed is faster than SIFT detector. But its performance under the perspective and light change is bad. Therefore, we use the ORB detector which was advanced byE Ruble on ICCV (IEEE International Conference on Computer Vision) in 2011. The ORB [4] (Oriented FAST and Rotated BRIEF) is based on the famous FAST feature detector and the BRIEF feature descriptor which was put forward in 2010.

3.1. Feature Point Detector

Fast [5] feature point detecting algorithm is based on image gray value of feature point and its neighborhood points. If there are enough points around the candidate point whosegray value isenough different fromthe candidate point, we can draw the conclusion that the candidate point is a feature point.

$$\mathrm{N} = \sum_{x\forall(circle(p))} |I(x) - I(p)| > \varepsilon_d \text{(1)}$$

I (x) is the gray value of an arbitrary point on a circle around the candidate, I (p) is gray value of center point of the circle, and N represents the number of the points from the central point's neighborhood whose difference of gray level between itself and the candidate point is bigger than the set threshold. Generally speaking, if three-quarters of the points around the candidate point meet the requirement, we consider that p is a feature point.

However, Fast detector exists drawbacks: 1) Fast detector do not have corner response function, and it will get a lot of edge points; (2) Fast detector does not produce multi-scale feature. (3) Feature points detected by Fast detector do not have the orientation information. Aiming at removing these defects, ORBmethod has advanced some improvements: 1) if you want to get N key points, you should reduce the threshold value to get more than N key points, then order them according to the Harris measure, and pick the top N points. 2) We use a scale pyramid of image to get feature points in each level. 3) The orientation of feature points can be calculated by intensity centroid which is shown below:

$$M_{ij} = \sum_x \sum_y x^i y^j I(x,y) \qquad (2)$$

$$C_x = \frac{M_{10}}{M_{00}} \qquad (3)$$

$$C_x = \frac{M_{01}}{M_{00}} \qquad (4)$$

$$C_{ori} = tan^{-1}(\frac{C_y}{C_x}) \text{(5)}$$

C_x Is offset of intensity from its center in x coordinate,$C_y$is offset of intensity from its center in y coordinate. We consider the C_{ori} as the orientation of feature point.

3.2. Feature Point Descriptor

ORB method uses BRIEF [6] descriptor to describe the keypoint. It first picks a small patch around the feature point from the smoothed image, and then obtains a bit string by testing point pairs which are chosen from the points around feature point's neighborhood. The bit value is determined by the mathematical formula below:

$$\tau(P;x,y) = \begin{cases} 1: P(x) < P(y) \\ 0: P(x) \geq P(y) \end{cases} \quad (6)$$

$P(x)$ is the gray value of point x and $P(y)$ is the gray value of point y. So the descriptor could be represented by a vector whose length is n (n is set to 256 generally). As the below show, the value of i can be 1 to 256 meaning different bit. The value on different bit can be 1 or 0, thus it composes a bit string whose length is 256.

$$f_n(P) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(P; x_i, y_j) (7)$$

In order to make BRIEF descriptor be invariant to rotation, ORB has improved the BRIEF to steered BRIEF. This method first makes up a matrix by the testing point pairs,

$$\mathrm{S} = \begin{pmatrix} x_1 & \cdots & x_n \\ y_1 & \cdots & y_n \end{pmatrix} \quad (8)$$

Then gets the rotation matrix $R_\theta$with the angleθ of feature point, and makes up a $S_\theta$matrix below

$$S_\theta = R_\theta S \ (9)$$

At last gets steered BRIEF descriptor:

$$g_n(P,\theta) := f_n(P) \mid (x_i, y_i) \in S_\theta \quad (10)$$

When we obtain the steered BRIEFdescriptor, the testing point pairs will have large correlation and low variance. To solve these problems, ORB develops a learning method for choosing good testing point pairs and this method is greedy search. The improved steered BRIEF can be called rBRIEF. Experiments have proved it has a good performance.

3.3. Feature Point Matching

Feature points matching and database query and image retrieval is essentially the same question, only a little different on the data size. They can be summed up by searching the nearest one in high dimensional vector with a distance function of similarity, so how to design the data index structure will greatly affect the efficiency of searching. KD [7] tree (k - dimension tree for short) is a kind of data structure in K dimension data segmentation. It is mainly used to search key data in multi-dimensional space (such as: range search and nearest neighbor search). However, if the data is a higher dimensional vector, the search efficiency of standard KD searching is extremely low. Practice have proved that the data dimension of standard KD tree search should not larger than 20, which limits the use of KD tree searching. Therefore, KD tree searching should be improved to the BBF (Best Bin First) [8] search method. There is no difference between KD tree search and BBF in nature. It mainly takes into account of the priority concept and shows good properties in the high dimensional vector search. Due to the ORB feature descriptor is 256 D vector, we use the BBF method to match points. At the same time, we use hamming distance to measure the similarity between points. In information theory, the hamming distance between two strings is the number of the different characters of two strings in corresponding position. In other words, it is the number of characters that need to be replaced to become the other string. Here it means the difference on the corresponding 256 bits of data. If the feature points are similar, the distance is short. The otherwise, it is far.

The inevitable situation is that mismatching points exist. In order to eliminate the influence of wrong matching, a cross - check method is adopted, and namely the data in the matching points must be the nearest data in its own data set. If the matching points do not meet the requirement, the matching will be removed. It can eliminate the mismatching roughly to make the matching points correct as much as possible.

4. Transformation Matrix Acquisition Algorithm

In the above section, we have obtained matching points between images. Thus we can obtain approximate transformation matrix between two images according to the matching points. As it is shown in Figure 1. A is a real point in 3 D space whose projective point on imaging plane $o_l$and $o_r$is$a_l$anda_r. These projective points satisfy the following formula:

$$\begin{pmatrix} u_l \\ v_l \\ 1 \end{pmatrix} = P1_{3\times4} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad (11)$$

$$\begin{pmatrix} u_r \\ v_r \\ 1 \end{pmatrix} = P2_{3\times4} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad (12)$$

Thus, transformation relations between the two-dimensional point of a_l and $a_r$is a 3 x3 matrix:

$$\begin{pmatrix} u_l \\ v_l \\ 1 \end{pmatrix} = M_{3\times3} \begin{pmatrix} u_r \\ v_r \\ 1 \end{pmatrix} (13)$$

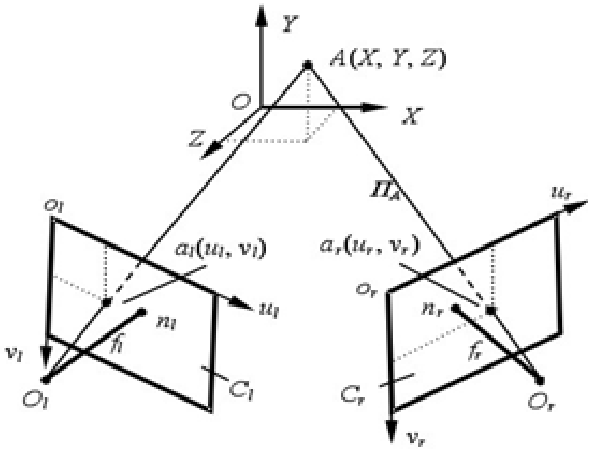


Figure 2. Object Point's Projection on Two Imaging Plane

We've got a lot of matching feature point pairs. In order to calculate the transformation matrix, we only need 4 pairs of matching points. This is an over-determined problem in numerical analysis. Aiming at solve the over-determined problem, we adopted the method of RANSAC (random sampling consensus method)to calculate transformation matrix to satisfy most matching points. In fact, problem is far more than that. The matching points we have got contain many mismatching points, so we need to remove these points as far as possible. According to the epipolar geometric constraint [9], a pair of matching point must satisfy the following relation:

$$(u_l \quad v_l \quad 1)F_{3\times3}\begin{pmatrix}u_r\\v_r\end{pmatrix} = \quad (14)$$

Bring a large number of matching feature points to the formula (12), we can obtain such a fundamental matrix F according to the RANSAC method. In turn, according to fundamental matrix F, we will test matching points to see whether they meet constraints, then we can remove large amount of mismatching points and get more accurate matching points finally. At last, bring the relative true matching points into the transformation matrix, we can get more accurate image transformation matrix.

5. Feature Points Searching in Small Region Algorithm

After the detecting and matching of feature points, there must be a lot of key points which have not been found automatically on target object contour. We can choose these key points manually on an image through the human-computer interaction forms, and then map them to the corresponding another image according to the transformation matrix calculated in the previous section. As various factors exist, the mapping points will not be exactly at the right location on the image. We can set a threshold value, namely make the selected point and mapping point as the center, take the threshold as radius to draw circles. In the two circles, we will detect and match points. After eliminating mismatching points, we select the matching points which have the least sum of distance with selected point and mapping point as the key point we want.

For the value of radius, we should consider the overall deviation degree of points in the image and search efficiency at the same time. For some larger deviation points, we can manually move it to a permissible deviation range, which will reduce the amount of calculation and improve the efficiency of search on the whole. The whole algorithm flow chart is shown below:

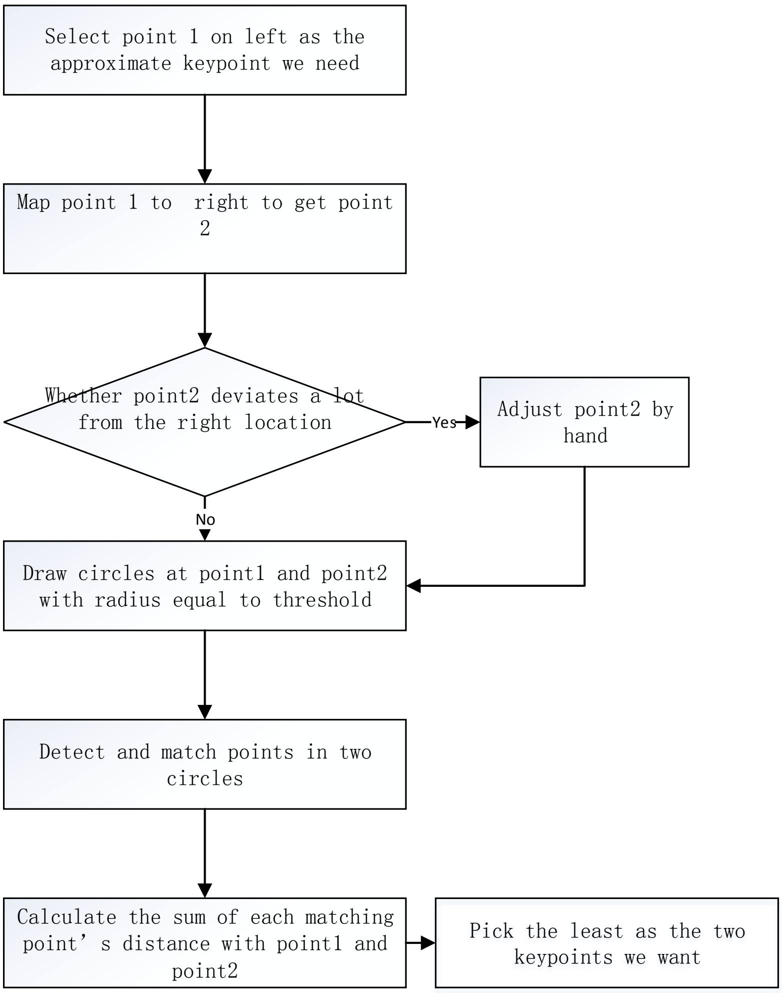

Select point 1 on left as the approximate keypoint we need
Map point 1 to right to get point 2
Whether point2 deviates a lot from the right location
Yes
Adjust point2 by hand
No
Draw circles at point1 and point2 with radius equal to threshold
Detect and match points in two circles
Calculate the sum of each matching point' s distance with point1 and point2
Pick the least as the two keypoints we want

Figure 3. Flow Chart of Algorithm

In order to realize the automatic extraction and matching of feature points in the two circles, we choose ORB method used in the first section to detect and match points. ORB method uses FAST detector which should set the appropriate parameter to control the number of feature points. In formula (1), the value of $\varepsilon_d$affects number of feature points, the ORB method set 20 by default. Due to characteristic of FAST detector that it has no corner response function, it has large responses along edge, so Harris corner measure has been used to order the Fast feature points. ORB method picks top 500 feature points as the final points by default. In order to control the number of feature points, we can not only set the threshold $\varepsilon_d$value but also set the number of feature points that we need. When search in small region, we could reduce the threshold to get more feature points, and then get the important feature points. Through a large number of experiments, we find that setting 10 as a threshold value can get an ideal result. At the same time, we can reduce the value of Harris corner response threshold. In this way, we can get more number of points, and avoid omitting key feature points due to low Harris corner response values.

6. Experiment and Conclusion

In order to verify the validity of the method, on the computer with the 3.20 GHz Intel Core, 4.00 GB memory, we use vs2010 to build a platform for 3 d reconstruction. First Input two images whose resolution are 481 x 740, then set 20 as

Fast operator threshold and set 1000 as upper limit value of number of points at the same time to control the number of automatic extraction points. The extraction effect as shown in the figure below:

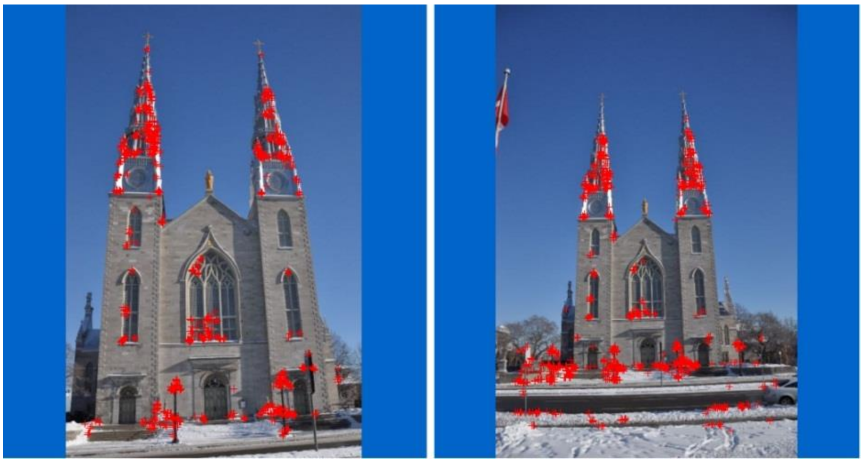


Figure 4. Automatically Extract Feature Point

It can be seen from the above that some key points in the image are not automatically discovered. We need choose them by hand at this time. For example, the two sharp corners on the building can’t be extracted. As is shown below, we can choose these two points manually on the left image, and make use of the transformation matrix relationship between two images to map the two points to right image.

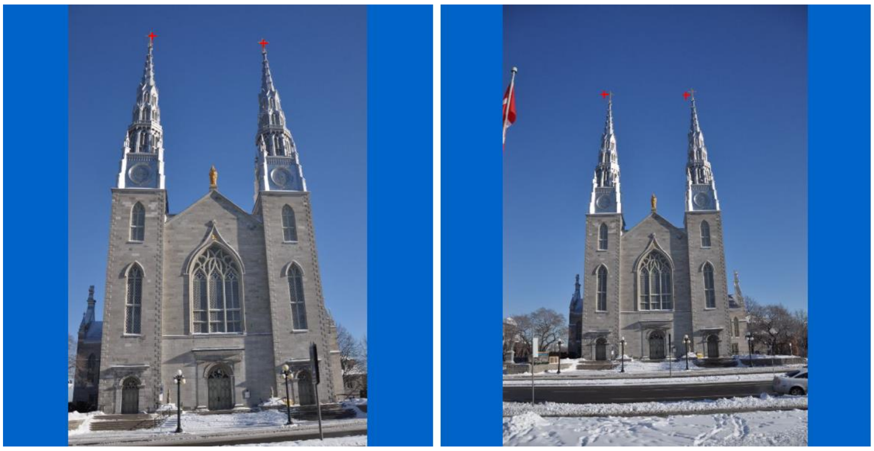


Figure 5. Select Key Points Manually and Mapping Key Points

It can be seen above that the selected points on the left have been mapped to the right image, but the mapped points on the right do not have exactly right locations. At these points, we respectively set threshold to draw circles to contain accurate locations, then detect and match points in these circles, at last achieve the desired key positions. The value of radius is determined by image resolution, the efficiency of search and neighborhood block size (usually for 31 x31) needed to calculate the descriptor. Here we set radius by 40 as area search scope. The experiment effect is shown below:

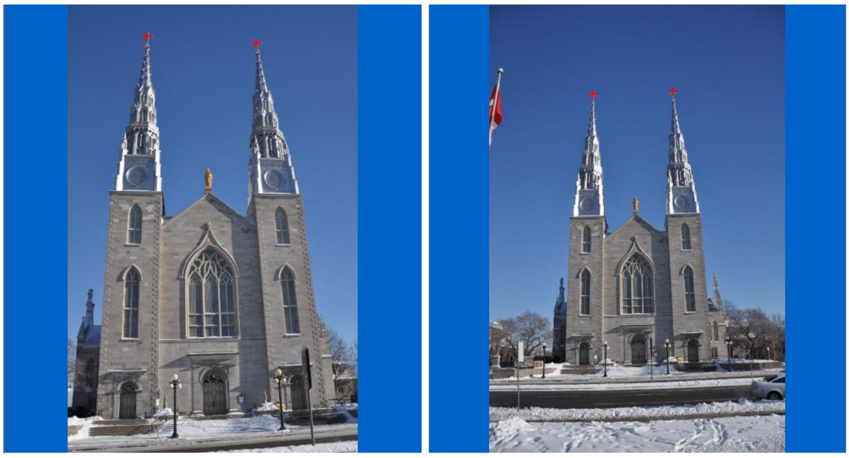


Figure 5. Accurate Key Position Searches

Select multiple feature points and statistical information as follows:

Table 1. Pixel Coordinate Statistics Pixel

N0	*Selected points on left*	*Mapped points on right*	*Adjusted points on left*	*Adjusted points on right*	*Left deviation*	*Right deviation*
point A	(173,158)	(134,61)	(173,157)	(133,59)	1	2.236
point B	(306,152)	(316,69)	(307,150)	(316,68)	2.236	1
point C	(246,327)	(237,284)	(246,326)	(237,284)	1	0
point D	(162,373)	(112,342)	(162,375)	(112,338)	2	4
point E	(325,370)	(352,348)	(328,374)	(353,347)	5	1.141
point F	(241,389)	(230,369)	(237,389)	(226,369)	4	4

From the table above, we can see that all the points we choose deviate the right positions more or less. Actually these deviations or residuals come from many aspects.

Firstly, in the stage of points matching, we use KD tree searching method to find the optimal matching points. Then we apply the cross-check method and epipolar constraint to remove the error matching points. However, all these processes cannot remove all error matching points. Therefore it will influence the accuracy of transformation matrix.

Secondly, in the stage of solving transformation matrix, we use the RANSAC (random sampling consensus) method to calculate transformation matrix to satisfy most matching points. Therefore the threshold that we use to evaluate whether the matching points are inliers also affect the accuracy of transformation matrix.

Thirdly, the points we choose by hand are scarcely possible on the right positions. So the mapped points we get from the products of selected points and transformation matrix will deviate a lot from the right positions.

At last, the points on the object may have different depth information. However, the transformation matrix we solve only meet the points whose depth information can be regarded as approximately equal. So if we select one point by hand whose depth is very different from others, the mapped points we get may deviates a lot from the right position.

Through the experiment, we can draw the conclusion that this algorithm based on man-machine coordination can solve problems we encounter in the 3 d reconstruction. The ORB algorithm used in the method is more quick and efficient than the traditional sift and surf. At the same time, searching in small region meets the requirement of finding important feature points and increases the searching efficiency.

Acknowledgements

This work is supported by the Special Scientific Research Fund of Public Welfare Profession of China (Grant No.201311001-1).I would like to express my truly gratitude to those who helped me during the writing of this thesis. First and foremost, I am greatly indebted to my supervisor, Professor Zhang Xiao Guo, for his patient instructions and suggestions on my thesis. I am also feel grateful to my fellows Peng De Qi, Ren Yan, Wan Xue Ying, and Wang Yun Fan.

Reference

[1] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision", Cambridge university press,, **(2003)**.

[2] D. G. Lowe, "J. Distinctive image features from scale-invariant key points", International journal of computer vision, vol. 2, no. 60, **(2004)**.

[3] H. Bay, A. Ess and T. Tuytelaars, "J. Speeded-up robust features (SURF)", Computer vision and image understanding, vol. 3, no. 110, **(2008)**.

[4] E. Rublee, V. Rabaud and K. Konolige, "ORB: an efficient alternative to SIFT or SURF", Proceedings of the 9th International Conference on Computer Vision, **(2001)**; Vancouver, Canada.

[5] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection", Proceedings of the 9th European Conference on Computer Vision, **(2006)**; Graz, Austria.

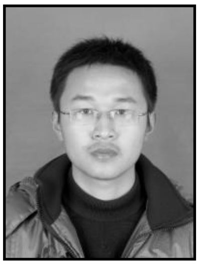
[6] M. Calonder, V. Lepetit and C. Strecha, "Brief: Binary robust independent elementary features", Proceedings of the 11th European Conference on Computer Vision, **(2010)**; Crete, Greece.

[7] J. L. Bentley, "J. Multidimensional binary search trees used for associative searching", Communications of the ACM, vol. 9, no. 18, **(1975)**.

[8] D. G. Lowe, "Object recognition from local scale-invariant features", Proceedings of the 7th International Conference on Computer Vision, **(1999)**; Kerkira, Corfu, Greece.

[9] Z. Zhang, R. Deriche and Faugeras, "J.A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry", Artificial intelligence, vol. 1, no. 78, **(1995)**.

Authors



Zhu Shi Wei, Master of mechanical and electronic engineering in Southeast University, Nanjing, China.

Zhang Xiao Guo, Dr. Xiaoguo Zhang is now with the School of Instrument Science and Engineering, Southeast University, as an associate professor. He received the M.S. and Ph.D. degrees in the CAD and GPS/DR/MM integrated vehicle navigation field from Southeast University, Nanjing, China, in 1998 and 2001, respectively. From 2002 to 2003, he was with the Maritime Research Center, Nan yang Technological University, Singapore.

His current research fields are focused on image processing, GIS, and 3S-integration.

Wang Qing was born in 1962 in China. He got his Ph.D. degree in the Department of Instrument Science and Technology from Southeast University, China, in 1998. Now he is a professor in the Department of Instrument Science and Engineering, Southeast University, China. His current research field is GPS/GIS theory and application.

Lv Jia Dong, was born in 1962 in China. He received the M.S degree in the School of Mechanical Engineering in Southeast University. Now he is a professor in the School of Mechanical Engineering in Southeast University in China. His current research field is electric light machinery and equipment.