DISCUSSION ON ROBOT EMBEDDED POSITIONING SYSTEM BASED ON IMAGE CORRESPONDENCE ALGORITHM

Yongjun Zhang
Hunan Urban Construction College, Hunan, China
Email: zhangyongjunurban@126.com

Abstract: In order to overcome difficulties in a structured environment and make image correspondence more efficient in image correspondence, and to realize a more robust, precise and efficient correspondence results that are suitable for complex scenes, Hough spectrum characterization and image contour is presented in the discussion. It is a novel and robust approach and it proves the features of Hough spectrum: translation, rotation zoom, and cycle; based on Fourier domain phase related algorithm, an efficient and robust sampling Fourier image correspondence is proposed; at the same time, estimate method based on monocular vision of motion robot is proposed in terms of the application of image correspondence in the motion estimation and positioning system. The experimental results show that this method is able to estimate the motion parameters with high precision, and eventually works out a reliable estimation of robot trajectory.

Keywords: Image correspondence; phase correlation; motion estimation; positioning system

1. Introduction

Image correspondence is one of the basic technologies in the field of image processing and machine vision. The purpose of image correspondence is to estimate the transformation parameters between two or more images to be matched to achieve the best alignment in space [1]. The typical application of image correspondence in computer vision is the binocular vision stereo correspondence, monocular image sequence optical flow, but also can be used for remote sensing, image correspondence in medical image processing, video compression, image super resolution image motion estimation and other fields, it can also be applied to many other areas: scene-together, virtual scene reconstruction, information fusion, pattern recognition, image correspondence, etc. The usual matching problem can also be extended to three-view and be generalized to the broader N- view example [2-4].

With the development of high precision correspondence requirements, scholars began to study how to apply certain correspondence problem with high accuracy and more robustness, and suitable for complex scenes. Such as global stereo matching with high precision method based on functions of cross validation, fitting method to estimate the parameter correspondence and optical flow estimation [5, 6], high precision based on image segmentation such as high precision sub-pixel rotation angle function of the angle error estimation.

This thesis aims at improving the accuracy of image correspondence algorithm, and overcome difficulties in a complicated environment, irregular texture effects and other problems. Some key problems of image correspondence technology based on signal phase were studied. These studies include: image correspondence of structured scenes and high precision of complex scene image correspondence and robot motion estimation. In-depth research made on image phase information based on image correspondence technology.

2. Methodology

2.1 Image correspondence technique

Image correspondence is a fundamental problem in the field of image processing and machine vision. It can be expressed in general mathematical form: given space $\mathbf{R}^d$, assuming that $V(x)$ is the pixel value of T in D dimension space $X = (x_1, x_2, x_3, \ldots x_d)$, then we use $\mathbf{I}(d) := \{V : \mathbf{R}^d \rightarrow \mathbf{R}\}$ to express the image of dimension. When $\mathbf{g} : \mathbf{R}^d \rightarrow \mathbf{R}^d$ represents a geometric transformation, then $S_g(x) : = S(x) \, g = S(g(x))$ is the image after g geometric transformation of the image S. Assuming that $\varphi : \mathbf{R} \rightarrow \mathbf{R}$ is pixel correction function, then $\varphi(S_g(x)) = \varphi \, S_g(x)$ represents the image obtained from pixel correction after image $S_g(x)$, and image correspondence is defined as:

The given distance function D: $I(d)^2 \rightarrow R$, and two images f, $s \in I(d)$, geometric transformation group G:$\{\mathbf{g} : \mathbf{R}^d \rightarrow \mathbf{R}^d\}$ and mapping function group ψ, the

image correspondence process is to find the transformation function g to shorten the distance between f and s (image corrected and yet to be matched), as shown below (1).

$$g = \arg\min_{\varphi \in \psi, g \in G} \left\{ D\left(f, g \cdot s \cdot \varphi\right) \right\} \qquad (1)$$

Image correspondence methods include: direct correspondence methods in time domain (such as correlation method, time-domain image feature method, mutual information method, etc.) and correspondence methods in transform domain (such as phase correlation method, phase difference method, etc.) [7].
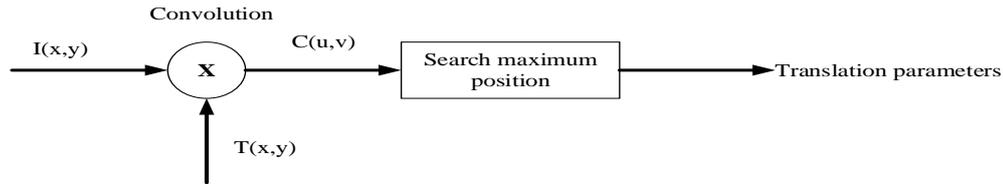
The image correspondence method in the time domain directly matches the image by the brightness information in the time domain. Cross correlation is a basic statistical method for image correspondence in time domain [8]. It can get the least squares solution of image correspondence problem and the correspondence process is shown in Figure 1.

Figure 1 Cross correlation correspondence flow chart

According to the properties of Fourier transform, the properties of the image translation, rotation, zoom and others have a corresponding response in the frequency domain, the translation of the two images correspond to the image phase difference. Image f_1 and f_2 fits $f_2(x,y)=f_1(x-d_x,y-d_y)$, then the below formula (2) is the phase correlation spectrum, in which $F_1(u,v)$ and $F_2(u,v)$ are the Fourier transform of $f_1(x,y)$ and $f_2(x,y)$. (*) denotes complex conjugate, (2) the inverse Fourier transform is a delta function with translation parameters as the centre. The inverse Fourier transform (IFFT) of equation (2) produces phase correlation spectrum, and the corresponding point to the maximum value is the parameters of translation for estimation.

$$R = \frac{F_1(u,v)F_2^*(u,v)}{\left|F_1(u,v)F_2^*(u,v)\right|} = e^{(ud_x+vd_y)} \qquad (2)$$

Time frequency transformation is a basic tool for signal processing and analysis. It decomposes the time domain signal into amplitude and phase in frequency domain. Amplitude and phase play different roles in signal reconstruction. A lot of important features of signals are stored in the phase [9]. When two images f (x, y) and g (x, y) can have translation motion (x_0, y_0), which is $f(x,y)=g(x-x_0,y-y_0)$, then F(u,v) and G(u,v) respectively are the Fourier transformation of the two images. According to the Fourier translation theory, the response of the translational motion in the frequency domain from the airspace is the linear phase difference of two signals in the Fourier domain, then F(u,v) and G(u,v) follows the below expression:

$$F(u,v) = G(u,v)\exp\left(-i2\pi\left(\frac{ux_0}{M} + \frac{vy_0}{N}\right)\right) \qquad (3)$$

The phase correlation theory is presented on the basis of the above formula, as shown in formula (4):

$$R_{fg}(x_0, y_0) = \frac{F(u,v)G^*(u,v)}{\left|F(u,v)G^*(u,v)\right|} = \exp\left(-i2\pi\left(\frac{ux_0}{M} + \frac{vy_0}{N}\right)\right) \qquad (4)$$

Among them, (*) denotes complex conjugate. $R_{fg}(x_0, y_0)$ is called the phase correlation spectrum, phase spectrum is the phase difference of two images, the phase correlation spectrum is made by inversing Fourier transform phase correlation spectrum response function. (x_0, y_0), is the centre of 2D Delta function $\delta(x-x_0,y-y_0)$, search the position corresponding to maximum response function, namely the translation parameters to be estimated, estimation process as shown in figure 2.
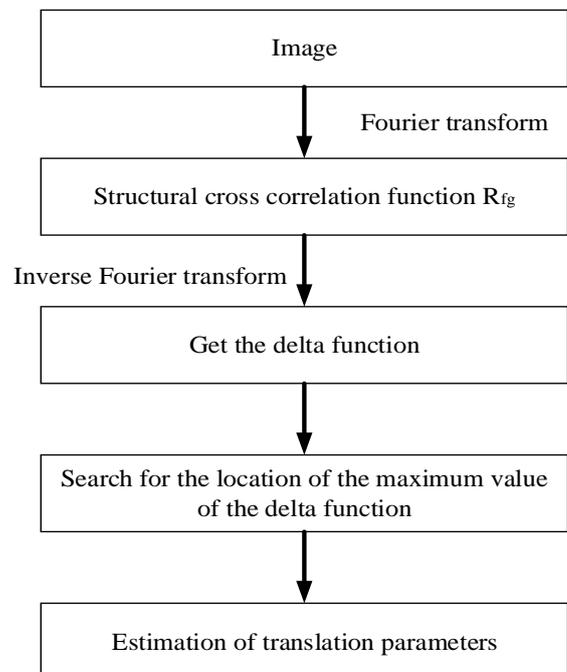

Figure 2 Phase correlation method to estimate the motion of image translation

2.2 Image correspondence methods

Hough transform is usually used to detect straight lines, circles and other analytic geometry in an image. It has phase characteristics. The advantage of Hough transform is to transform the more difficult global detection problem into the relatively easy local peak detection problem in parameter space. In addition, Hough transform is also one of the classical parameter estimation methods, but the huge space complexity and time complexity become unfavourable factors in its application. Therefore, Hough transform is mostly carried out in the feature space of image [10,11].

The point set P in image space S, after the Hough transform obtains the corresponding HT[θ,ρ]; the given function g features translational invariance, i.e. given f'(t)=f(t+a), then we have g[f] =g[f']. In order to retain as much information of the image, esp. the shape and scale, as possible, and to facilitate the use of correlation function of local maxima to estimate the rotation angle φ. By referring to the related literature with translational invariance of the Hough spectrum, Hough spectrum with such feature is briefly defined in the paper, namely, it not only considers the parameter space division number, but also considers the radial pixel in the distance in Hough space. It is necessary for estimating the parameters for zooming.

$$HS_g[i](\theta) := \sum_{j=1}^{n} t_j (\rho_j - \bar{\rho})^2 \cdot \left(\sum_{j=1}^{n} t_j \right)^{-1} \quad (5)$$

The transverse coordinates in the Hough spectrum, θ, represent the spectral phase, and the ordinate $HS_g[i](\theta)$ represents the spectral amplitude. In the formula, t_j for all pixels by Hough transform in (J, 0) the number of voting, which is shown below

$$\bar{\rho} = \sum_{j=1}^{n} (\rho_j \cdot t_j) \cdot \left(\sum_{j=1}^{n} t_j \right)^{-1} \quad (6)$$

The correspondence Hough spectrum inherits the rotation and periodic characteristics of Hough space. It transforms the rotation into the translation of the spectrum in the phase axis, and transforms the scaling parameter into the spectrum to zoom along the amplitude axis. In addition, the matched Hough spectrum is invariant under translational transformation.

Fourier transform is one of the commonly used transformation methods in the field of signal processing. The theory of image registration based on Fourier transform is the theory of Fourier translation. However, the traditional domain Fourier phase correlation technique can only estimate parameters pixel, or as a more accurate method of initial estimates, the registration accuracy needs to be improved, sub-pixel registration method is proposed on the basis of the main zero fill (zero padding) method, phase correlation spectrum interpolation method and phase correlation spectrum the fitting method, high precision [12] to achieve high efficiency.

In general, for two-dimensional discrete signal f (x, y), its Fourier transformation is defined as formula (7), of which u=0, 1... M-1, v=0, 1... N-1.

$$F(u,v) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \exp\left[-j2\pi \left(\frac{ux}{M} + \frac{vy}{N} \right) \right] \quad (7)$$

The Fourier translation theory is the basis of image registration in the Fourier domain, that is, the translation of signals in the time domain corresponds to their linear phase difference in the frequency domain. The traditional POC (phase correlation) technology can only estimate the parameters of pixel level precision, and its registration accuracy needs to be improved. In this paper, an Enhanced Phase Only Correlation (EPOC) method based on frequency domain local sampling Fourier transform is proposed, which can accurately process image correspondence [13]. EPOC constructs a matrix based up sampling transform in the local region around the peak of pixel level accuracy of phase correlation spectrum, and obtains robust, efficient and accurate sub-pixel estimation parameters. The classical POC method as the pre-treatment of EPOC POC, the estimated phase correlation peak as the construction of sampling matrix parameters, and the inverse Fourier transform matrix in the form of a small neighbourhood around it, get the Delta function, so as to obtain sub-pixel precision registration parameters. The flowchart of EPOC for image translation estimation is shown in Figure 3.
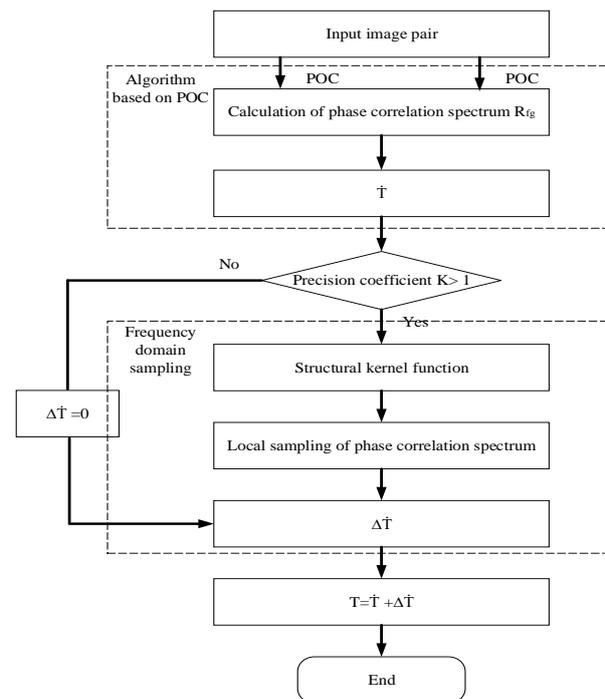

Figure 3 Flow chart of image translation estimation based on EPOC

The analysis of time-frequency transformation indicates that the translation of image only affects the phase of the image in the frequency domain. The rotation and scaling parameters of the image are consistent with the rotation and scaling of the amplitude spectrum of the image in the frequency domain, as shown in Figure 4.

Therefore, the image in the Cartesian coordinate system rotation and zoom can be transformed into the translation in the amplitude spectrum in the log-polar space, a phenomenon known as the frequency response of rotation and zoom phenomenon, namely amplitude spectrum is invariant in translation, when the signal is transformed into frequency domain when rotating, zoom and pan are decoupled.
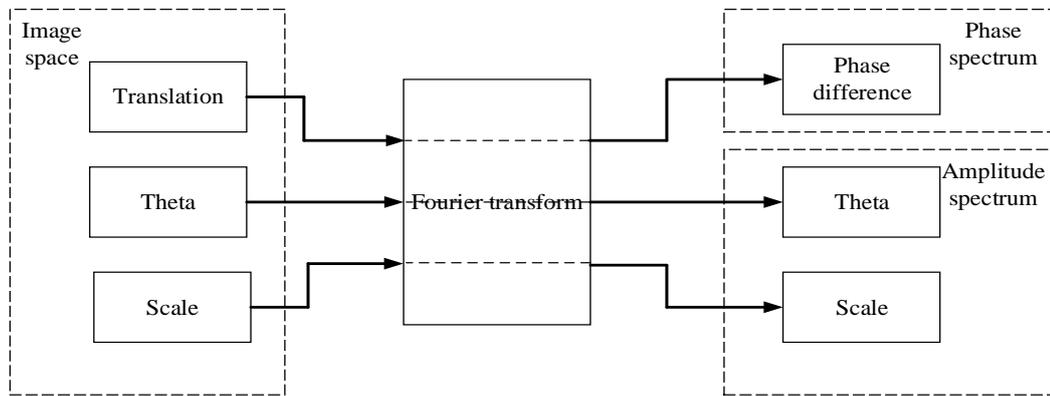

Figure 4 Fourier domain parameter decoupling

2.3 Image motion and robot motion model

The system architecture of estimating the motion parameters of the robot with a single camera as a sensor is shown in Figure 5. The system consists of two main parts: the motion estimation of the image and the estimation of the motion parameters. The corrected image sequence is used as the input of the motion estimation part of the image. The motion parameters are calculated with multi-scale optical flow field as input parameters, and the yaw angle, pitch angle and translation parameters are estimated according to the robot motion model and the block RANSAC algorithm. Motion estimation system shows the process of estimating motion parameters of platform from monocular vision image sequence, which includes two main steps: image correspondence, instantaneous relative motion measurement in image space and instantaneous relative motion measurement in 3D world coordinates.
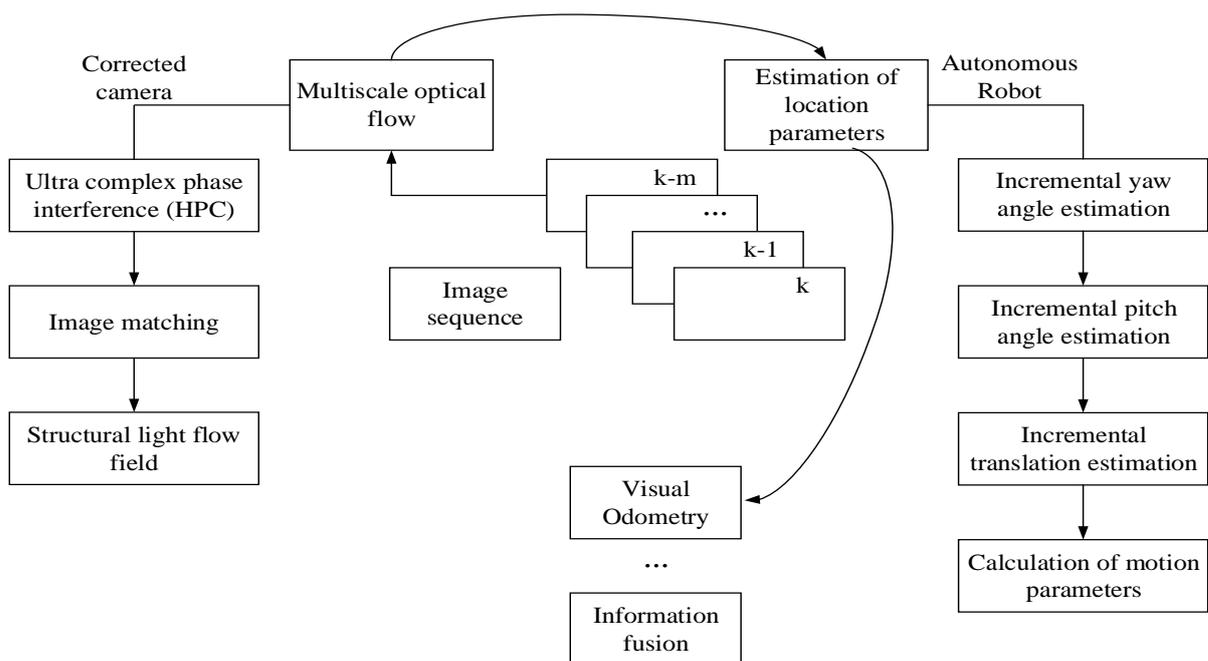

Figure 5 Motion estimation systems for monocular vision robot

Assuming that the robot is in a static environment, the coordinate system OXYZ is fixed to the forward camera. As shown in Figure 6, the Z axis is in accordance with the optical axis. The rigid body motion between the space points can be decomposed into rotational motion $\mathbf{\Omega}= (\Omega_X, \Omega_Y, \Omega_Z)^T$ and the translational $\mathbf{T}= (T_X, T_Y, T_Z)^T$. The coordinates of the scene point P in the camera coordinate system are represented by (X, Y, Z). (x, y) represents the corresponding pixel coordinates on the image plane, and the image plane is located at the focus, namely $Z=f_c$. The motion of the camera can also be considered to be the motion of the static scene point P relative to the camera coordinate system.
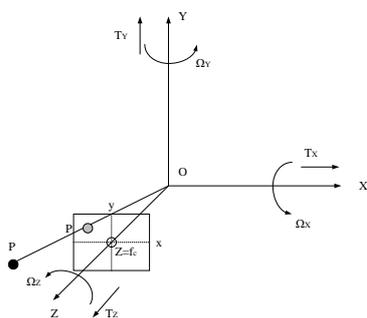

Figure 6 Coordinate system

By the perspective geometry relation, the motion of the two-dimensional image point (x, y) in the image plane (U, V) is expressed as

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \dfrac{-f_c T_X + x T_z}{z} + \dfrac{xy}{f_c}\Omega_X - \left(\dfrac{x^2}{f_c} + f_c\right)\Omega_Y + y\Omega_Z \\ \dfrac{-f_c T_Y + y T_z}{z} + \left(\dfrac{y^2}{f_c} + f_c\right)\Omega_X - \dfrac{xy}{f_c}\Omega_Y - x\Omega_Z \end{bmatrix}$$

(8)

The above formula features the model of the relationship between the motion of the robot and the motion of the image. The influence of robot translation and rotation on image production is independent and separable, and the rotation components of motion are independent of scene structure. Scene depth Z only affects translation components.

The visual odometer consists of two main parts: ego-motion estimation and motion trajectory reconstruction. The former estimates the incremental motion according to the optical flow field, and the latter restores the position and attitude of the robot in the world coordinate system. 4-DOF self-localization model based on incremental estimation of separate rotation and translation parameters, and rotation parameters can be estimated by the optical flow vector of far field, when the pitch angle and yaw angle compensation, the optical flow vector of the area near to robust estimation of translation parameters, Figure 7 is the image of the optical flow image and the far zone (far region) and near (near region) schematic partition.
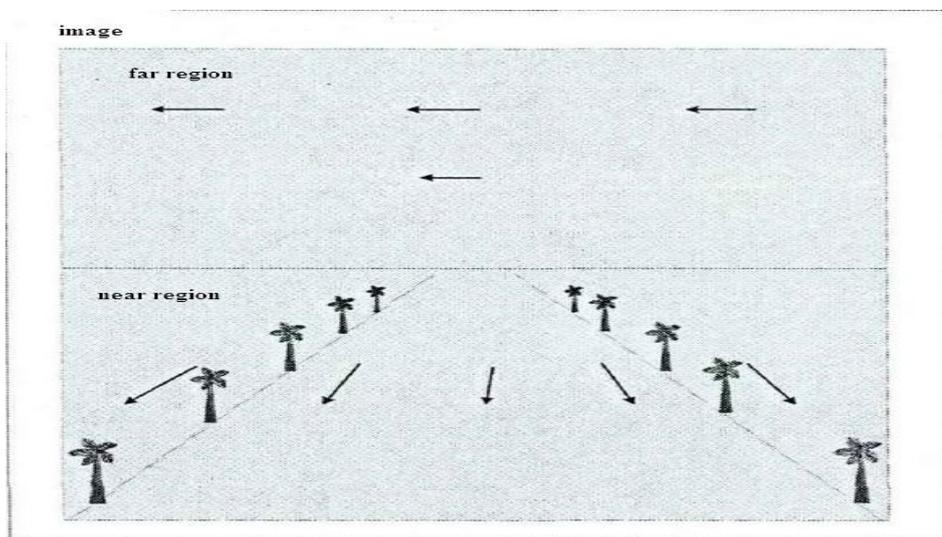

Figure 7 Image partition schematic diagram

3. Results and Discussion

3.1 Hough spectrum related image registration experiment

The correspondence process of Hough spectrum image registration, the first two groups of shape boundary coordinate data is mapped to the Hough space. Construct correspondence Hough spectra of the two Hough spectrum to process, and obtain the translation angle φ, i.e. the rotation angle between shapes; estimate shape zoom factor α; and process related Hough density function after alignment, calculate the shape shift T. Through this series of operations, the registration between two shapes is completed, that is, the estimation of the rotation,

zoom and translation parameters. The detailed flow chart is shown in Figure 8. The key steps of registration algorithm include: the construction of Hough spectral correlation function, the extraction of

rotation parameter φ, the construction of Hough spectral scaling function and the scaling factor α, the construction of Hough density correlation function and the extraction of translational parameter T.
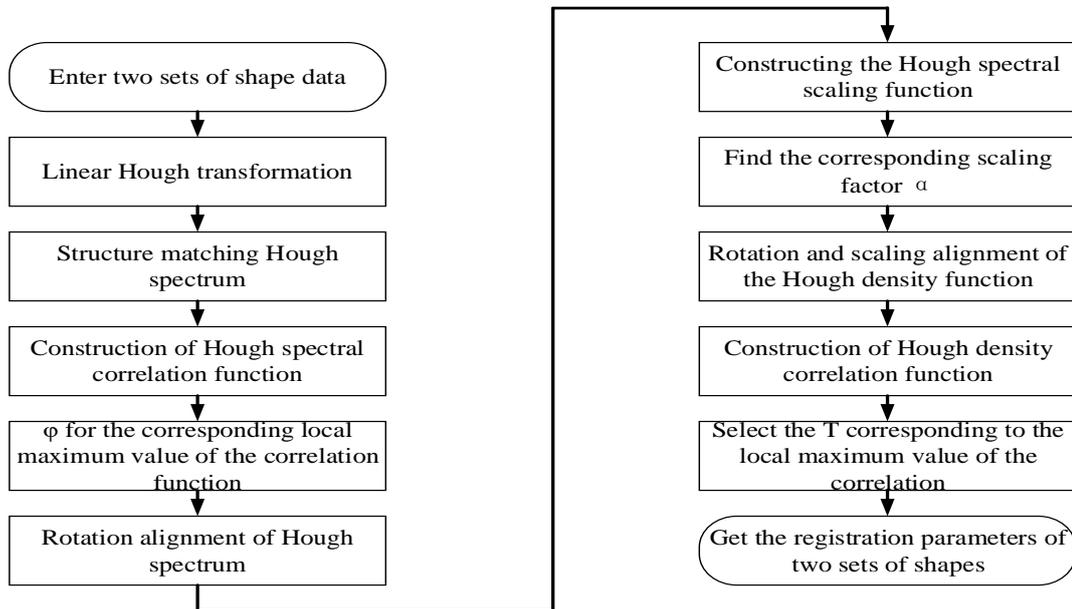

Figure 8 Shape registration flow chart of Hough spectrum

Features including translation, rotation, scaling and periodic, makes Hough spectrum suitable for image registration. In order to verify its performance, a lot of experiments are carried out on the shape provided by SIID shape set. Figure 9 corresponds to the shape of the Hough spectral representation shows the estimation results of shape registration parameters based on Hough spectrum of the registration results is to shape registration plans in

accordance with the registration results are rotating, zoom, scale and shift transform and template shape graph display results in the same coordinate system. The upper graph shows the image of the registration parameters. The experimental results show that the registration method based on Hough spectrum has good registration effect and precision for any complex shape.
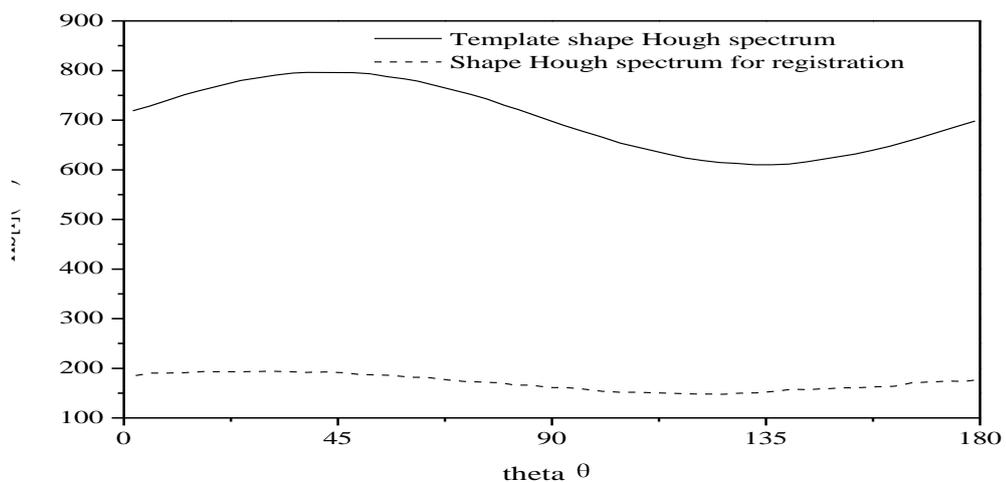

Figure 9 Hough spectrum representation of shape

In order to further verify the effect on the accuracy of registration, different levels of zero mean

white Gauss noise are added in the registration form in an image.

The actual translation, zoom, rotation parameters: (-15, -11), 1.4 pixels, 12 degrees, respectively, and

the shape of template matching results are shown in table 1.

Table 1 Registration results under different levels of noise

Group	Noise variance δ	Translation (pixel)	Scale	Theta parameter (degree)
1	0	(-15,-11)	1.4044	12
2	0.03	(-15,-11)	1.4020	12
3	0.07	(-15,-11)	1.3977	12
4	0.10	(-15,-11)	1.4018	12
5	0.12	(-18,-11)	1.3930	11

When adding proper amount of zero mean Gauss white noise to registration shapes (such as delta < 0.10), but sees no significant influence on translation, rotation of the registration parameters, then it has little effect on the zoom parameter. When the noise is stronger (such as delta =0.12), the noise has a great influence on the registration results. When the contour of a shape is extracted, the effect of noise can be further reduced if appropriate noise suppression algorithm is adopted. In general, the registration method based on Hough spectrum has strong ability to resist noise.

In order to verify the adaptability of the image correspondence method based on matching Hough spectrum registration, a comprehensive test of various shapes, various translation amplitudes, various scaling ratios and quadrant rotation angles is carried out on the SIID shape set. Put forward to the related literature of Hough spectrum is in the SIID data sets of matching experiments, figure 10 shows the experimental results of statistics, the experiment was divided into anti-noise performance and anti-blocking performance of two parts, anti-noise experiments, adding random variance in the range of 0~0.2 of the Gauss white noise in the image, and statistics experiment 50 times. In the anti-occlusion experiment, the size of the 0~1/5 image was randomly added to the occlusion and carried out 50 statistical experiments. The experiment shows that the method is better than the literature method in the field of anti-noise and occlusion. To sum up, the matching method based on matched Hough spectrum has strong adaptability and has important application value
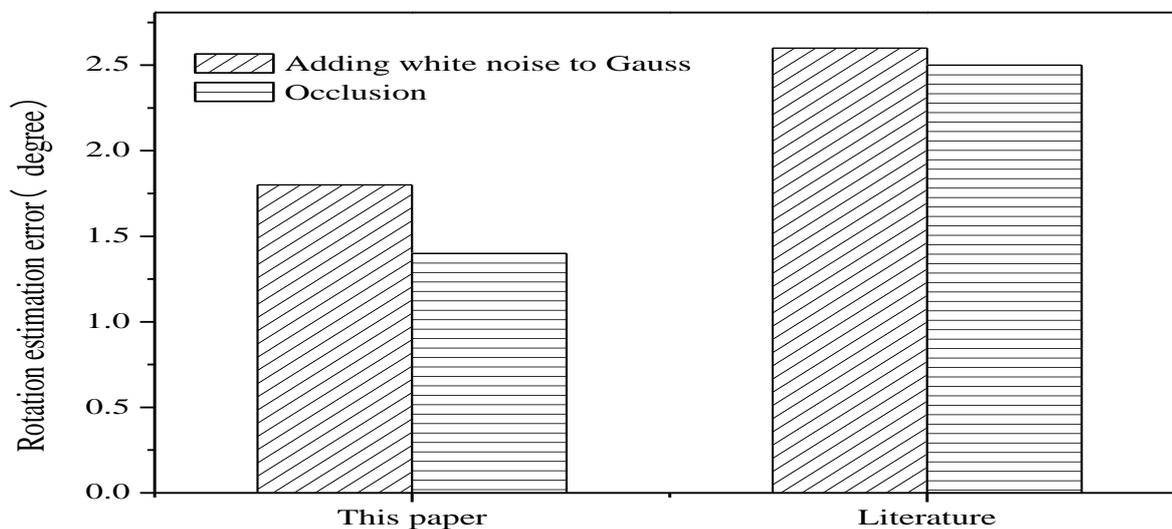

Figure 10 Comparison of experimental results on Camel images of SIID shape set

3.2 Performance and registration experiment of EPOC algorithm

The accuracy of the EPOC algorithm for rotation and translation estimation is evaluated through the classic sub pixel registration and the EPOC method. Table 2 shows the average error and maximum error statistics of the traditional POC method, the EPOC method and the PPFT method for estimating the

rotation angle. Through experiments and analysis, we find that when the accuracy coefficient is 10, the average error estimated by EPOC method is about 0.1 degree, and the maximum error is less than 0.8 degree on a given image. The experiment shows that EPOC method can estimate with a higher precision of rotation angle compared with PPFT method.

Table 2 Statistical error of rotation angle estimations (°)

	POC		A method based on PPFT				EPOC（k=10）	
			Ω		$Ω_N$(normalized)			
	Average error	Maximum error	Average error	Maximum error	Average error	Maximum error	Average error	Maximum error
a	0.82	1.49	0.68	1.28	0.73	1.14	0.097	0.493
b	0.88	1.31	0.72	1.77	0.72	1.29	0.101	0.616
c	0.87	1.49	0.90	1.44	0.87	1.20	0.100	0.767
d	0.86	1.30	0.83	2.12	0.80	1.66	0.101	0.766

Further, from the description of the EPOC method, we find that the precision coefficient k affects the accuracy of registration. Now we analyze a series of experiments to estimate the effect of k on the estimation error in the process of high accuracy rotation angle estimation. The experiment uses 256 x 256 Lena image. The statistical curve of the estimated error of the rotation angle estimation with the precision coefficient k from 1 to 20 is shown in Figure 11. When k increases from 1 to 5, the registration error decreases rapidly. However, when it is more than 20, the estimation error is balanced. In a word, when k is between 5 and 20, we can get a higher precision of rotation angle estimation.
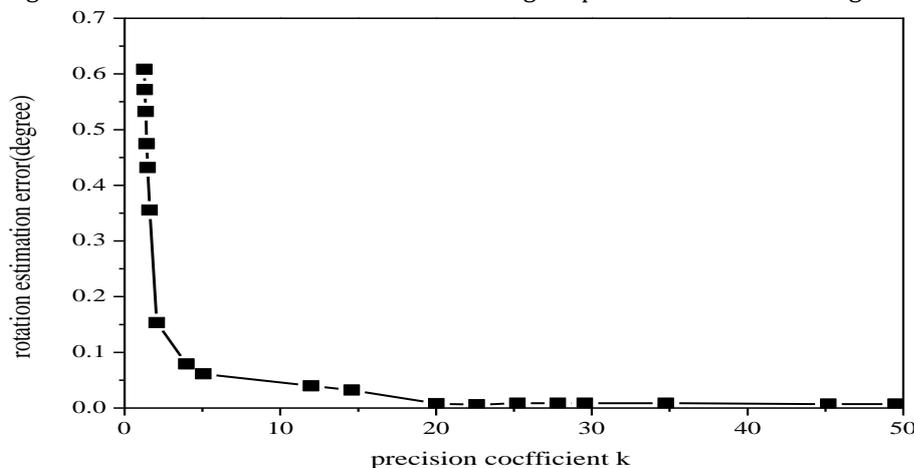

Figure 11 The influence of the accuracy coefficient K on the estimation error of the rotation angle

3.3 Experiment and analysis of robot location estimation based on image matching

The image sequence obtained by the independent platform is used as the data source to verify the mileage calculation method based on the monocular vision. From the example of running independent platform in rural road environment, continuous image from the platform is used to estimate the trajectory of the platform, analysis of the pitch angle increment can have pitch angle distribution and change trend of the entire process, it can be used as a quantitative road bumpiness measurement and it shows indirect characteristics of rural road that it is not smooth. The distribution of the pitch angle in the whole process is shown in Figure 12.
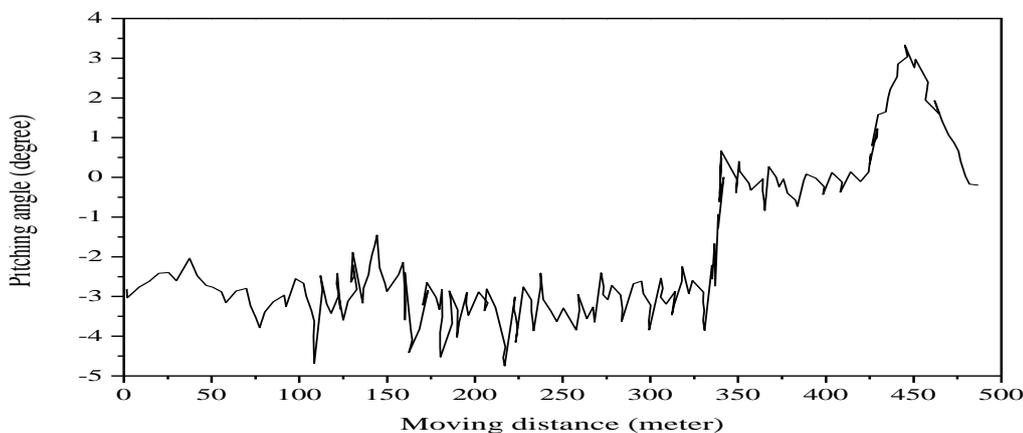

Figure 12 The distribution of the pitch angle increments relative to the marching mileage during the whole experiment

The 3DOF model, visual odometer, scale invariant feature (SIFT) based visual odometer method and phase based derived optical flow and 4DOF self-positioning model visual odometer are compared. In the SIFT based method, SIFT feature detection operator is used to detect continuous image features. RANSAC algorithm is used to exclude outliers and optimize the basic matrix (Fundamental Matrix) and the essential matrix (Essential Matrix), and estimate 6DOF parameters to build the robot trajectory. Figure 13 shows the experimental results of the contrast, and the experimental results show that we propose a more optimized strategy compared to the 3DOF model or the SIFT based approach.
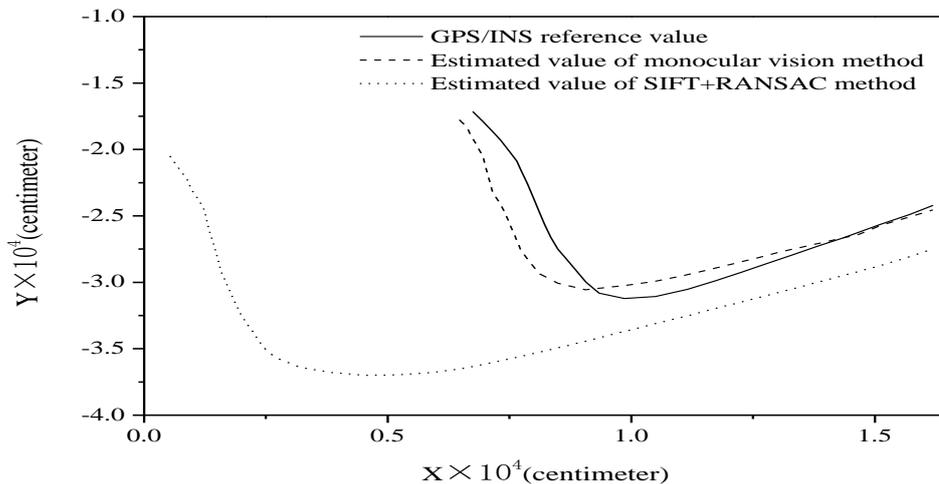

Figure 13 The trajectory of the estimation of the visual positioning system and its comparison with other methods

From the above experiments, we find that the optimized vision location system has broad prospects for application in combination with more robust and precise location and orientation systems. The system based on self-localization is super linearly accumulated along with the increase of robot's travelling distance. This inevitably restricts the application of the self-positioning method to the long-distance navigation system. However, if combined with other sensors, such as electronic compass, the long-distance performance of self-positioning system can be greatly improved, and the cumulative error can be reduced to a linear relationship with the operation distance.

It is well known that when the matching of signal level between two images is matched with the feature level, the former is very weak in computing efficiency. In fact, the time consumed for optical flow estimation induced by super complex wavelet phase is much more than that based on SIFT. The time consumption of RANSAC is mainly dependent on the number of optical flow vectors or features. In general, the proposed approach is slightly more time-consuming than the SIFT and RANSAC approach, and Figure 14 shows a comparison between the two in terms of average time consumption.
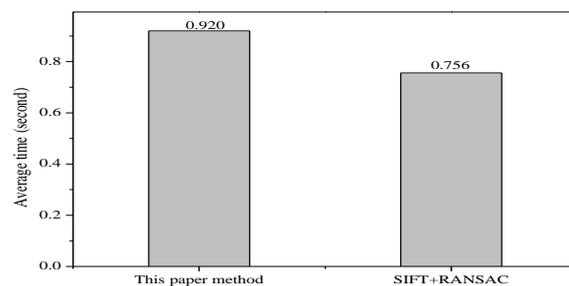

Figure 14 Comparison of average time consumption

4. Conclusion

In this paper, we propose a new algorithm based on image phase information, including structured image matching, high-precision image matching and phase-based image correspondence. By matching the Hough spectrum estimation of shape translation, rotation and zoom between the effective parameters of noise and occlusion, it has shown good robustness and high precision; for high precision image correspondence problem, phase correlation algorithm with an efficient and robust method of image sampling is proposed on the frequency domain. Through these algorithms, the accuracy and robustness of the image correspondence are improved and the applicability of the complex scene images, such as the robot positioning system, and so on.

Combined with the latest research progress in image correspondence technology, the new theory and new method of image matching need to be further considered. In addition, image correspondence and accurate estimation under the condition of nonlinear registration, such as noise or natural texture, is still a major and difficult in the future.

References

[1] Zhang Y R, Yang X, Qiao H, et al (2016). Introducing locally affine-invariance constraints into lunar surface image correspondence. Neurocomputing, 186: 258-270.

[2] Wang L, Li R, Fang Y (2016). Energy flow: image correspondence approximation for motion analysis. Optical Engineering, 55(4): 043109-043109.

[3] Remondino F, Spera M G, Nocerino E, et al (2014). State of the art in high density image matching. The Photogrammetric Record, 29(146): 144-166.

[4] Solsona S P, Maeder M, Tauler R, et al (2017). A new matching image preprocessing for image data fusion. Chemometrics and Intelligent Laboratory Systems, 164: 32-42.

[5] Wang Q, Shi W, Atkinson P M (2014). Sub-pixel mapping of remote sensing images based on radial basis function interpolation. ISPRS Journal of Photogrammetry and Remote Sensing, 92: 1-15.

[6] Plyer A, Le Besnerais G, Champagnat F (2016). Massively parallel Lucas Kanade optical flow for real-time video processing applications. Journal of Real-Time Image Processing, 11(4): 713-730.

[7] Kukkonen M, Maltamo M, Packalen P (2017). Image matching as a data source for forest inventory–Comparison of Semi-Global Matching and Next-Generation Automatic Terrain Extraction algorithms in a typical managed boreal forest environment. International Journal of Applied Earth Observation and Geoinformation, 60: 11-21.

[8] Sedaghat A, Ebadi H (2015). Remote sensing image matching based on adaptive binning SIFT descriptor. IEEE transactions on geoscience and remote sensing, 53(10): 5283-5293.

[9] Suen Y, Xiao S, Hao S, et al (2016). Time-frequency representation measurement based on temporal Fourier transformation. Optics Communications, 376: 86-91.

[10] Mukhopadhyay P, Chaudhuri B B (2015). A survey of Hough Transform. Pattern Recognition, 48(3): 993-1010.

[11] Cha Y J, You K, Choi W (2016). Vision-based detection of loosened bolts using the Hough transform and support vector machines. Automation in Construction, 71: 181-188.

[12] Baker M J, Trevisan J, Bassan P, et al (2014). Using Fourier transform IR spectroscopy to analyze biological materials. Nature protocols, 9(8): 1771-1791.

[13] Khyam M O, Ge S S, Li X, et al (2017). Highly Accurate Time-of-Flight Measurement Technique Based on Phase-Correlation for Ultrasonic Ranging. IEEE Sensors Journal, 17(2): 434-443.