Research on local visual global localization method based on out-of-view reference of spatial point association

Peipei Gao[1*], *Feng* Liu[2], *Qingping* Yang[3], and *Jiale* Wang[2]

[1] College of Computer Science, Nankai University, Tianjin 300071, China
[2] State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China
[3] College of Engineering Design and Physical Sciences, Brunel University of London, Uxbridge UB8 3PH, UK

Abstract. The global localisation of spatial points is a critical step in tasks such as object tracking, motion analysis and pose measurement. This paper addresses the critical issue of global localisation when spatial points are scattered and cannot be contained within the same field of view. It proposes a local visual global localisation method based on an out-of-view reference through spatial point association. By constructing a local measurement and localisation model using parallel binocular vision and a spatial coordinate transformation model that associates local regions with the global reference, the global localisation of spatial points inside and outside the field of view is achieved. Experimental results demonstrate that the localisation accuracy of spatial points is less than 0.1 mm in terms of distance measurement. This method is useful for cooperative multi-camera localization and multi-point measurement in large 3D spaces.

1 Introduction

In applications such as motion analysis, object tracking and size measurement, the extraction and localisation of spatial points is essential [1-2]. Spatial point localisation can be categorised into two distinct types: local and global. The focal point of this discourse pertains to the concept of global localisation in the context of out-of-view references.

A variety of methodologies are employed by disparate research domains to address this issue. In the domain of measurement, conventional methodologies employed for non-common field-of-view imaging scenarios encompass the utilisation of scene stitching to extend the visual field, and the employment of robotic arms to facilitate camera manipulation. For instance, grating projection combined with robotic arms or image stitching from multiple views has been employed [3-4]. The former utilises the pose parameters of the motion mechanisms to calculate the spatial points, while the latter achieves localisation by stitching images from different viewpoints or from multiple views. However, grating projection is not suitable for sharp elements or metal surface glare and is especially insensitive to small, sharp elements such as vertices and corners. The process of moving imaging measurement is inherently space-consuming, rendering it ill-suited for applications characterised by elevated spatial volume requirements. Moreover, it is important to note that the utilisation of these methodologies is not feasible in real time, as their operation entails the implementation of time-sharing measurements.

In the field of surveying, spatial coordinate transformation methods are employed to unify two distinct coordinate systems, with a focus on the transformation between systems that possess the same unit scale. The objective of this unification is to achieve localisation in a consistent manner. The objective of this study is to enhance conversion accuracy by optimising pose parameters based on public point cloud overlap. The following methods are typically used for coordinate transformation: the Bursa model, the Rodrigues matrix, the quaternion method, singular value decomposition (SVD) and the least squares method [5-7]. The establishment of mappings between coordinate systems [8] by these methods enables the localisation of points that are out of view. It is evident that methodologies such as Gauss–Newton and Levenberg–Marquardt utilise non-linear optimisation in order to ascertain the optimal pose transformation matrix.

The focal point of this paper is the localisation of points within the camera's field of view, and the subsequent association of these points with out-of-view global references. The primary challenges can be categorised as follows: firstly, local spatial points are visible but the global reference is not; secondly, space is limited, which prevents the use of mechanical guides; and thirdly, the spatial distance between local and global reference points must be calculated. The proposed methodology is a global localisation technique that utilises local parallel binocular stereo vision. Experiments have been designed to verify the method's effectiveness.

2 Problem Description

* Corresponding author: watersky@nankai.edu.cn

https://doi.org/10.1051/matecconf/202541301001

The entity object in this study is shown in Fig. 1, and based on the graphical resolution, the corresponding topology structure is generated as shown at the top of Fig 1. Dashed lines represent axes and denote invisible virtual elements, serving as localization references. The key information comprises spatial points 1 and 2, and central line reference element 3.

The distances between points 1 and line 3, points 2 and line 3, and between points 1 and 2 are critical geometric parameters. It is evident that spatial points 1 and line 3 are not amenable to contact measurement. The substantial span of these structures exceeds the direct field of view of optical systems, thereby rendering visual direct measurement unfeasible. The existing equipment is incapable of meeting the simultaneous requirements for real-time and high-precision localization. Consequently, the development of a rapid, precise and indirect global visual localization method is imperative.

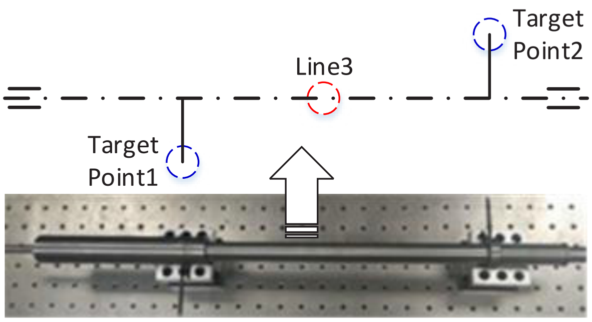

Fig.1. Schematic of physical structure diagram

3 Local Visual Global Localization Method Based on Out-of-View Reference of Spatial Point Association

The method under scrutiny here is one which targets the deviation of points 1 and 2 in the Z-axis depth direction, as well as their global spatial relationship. The prerequisite for implementation is the acquisition of complete positioning information for spatial points 1 and 2. The subsequent section will focus on the globalisation method of spatial points.

3.1 Measurement Principle

The system design layout is illustrated in Fig 2. The visual model employs a perspective imaging model. Due to spatial limitations, the two cameras are arranged in an upper-lower configuration, with close proximity to minimise the baseline distance. The field of view of the two cameras is limited to the local area in which the target point is located, thereby ensuring high measurement accuracy in the local small area. The measurement principle is predicated on the utilisation of parallel binocular vision in order to locate and measure target points in the local region. A transformation matrix is then established between the target points within the field of view and the reference points outside the field of view. The transformation matrix is utilised to establish the association between spatial points both within and without the field of view, in addition to performing geometric measurements.

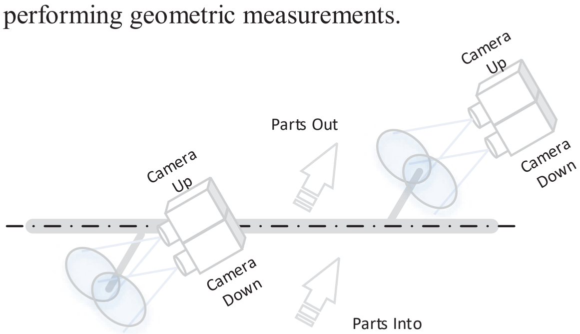

Fig.2. Layout of local optical-axis parallel binocular vision system

3.2 Target Point Localization in Local Region

Let P_w be the control point in the public area world coordinate system. If it appears within the overlapping area of the two cameras' fields of view, then the corresponding matching point U(u,v) can be found in the image of the upper camera. The transformation relationship vector between P_w and P_{up} is expressed as:

$$P_{up} = R \times P_w + T \tag{1}$$

In Formula (1), P_{up} represents the coordinates in the upper camera coordinate system, and P_w represents the control point coordinates in the world coordinate system. The expression represents a linear space transformation with reversible properties, where R and T describe the pose relationship between the camera coordinate system and the primary coordinate system. Rotation directions are defined by right-handed rotations around the X, Y, and Z axes, and the rotation matrix R can be expressed as in Formula (2); T is the translation vector:

$$\begin{aligned}
\boldsymbol{R} &= \boldsymbol{R}_z(\gamma)\boldsymbol{R}_y(\beta)\boldsymbol{R}_x(\alpha) \\
&= \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \\
&= \begin{bmatrix} \cos\beta\cos\gamma & \cos\alpha\sin\gamma + \sin\alpha\sin\beta\cos\gamma & \sin\alpha\sin\gamma - \cos\alpha\sin\beta\cos\gamma \\ -\cos\beta\sin\gamma & \cos\alpha\cos\gamma - \sin\alpha\sin\beta\sin\gamma & \sin\alpha\cos\gamma + \cos\alpha\sin\beta\sin\gamma \\ \sin\beta & -\sin\alpha\cos\beta & \cos\alpha\cos\beta \end{bmatrix}
\end{aligned} \tag{2}$$

Similarly, a similar equation applies to the lower camera. The pose between two coordinate systems includes 12 parameters:

$$\left[\begin{array}{l} Q_w \sim Q_{up}: (\alpha_{up1}, \beta_{up1}, \gamma_{up1}, t_{up1}, t_{up2}, t_{up3}) \\ Q_w \sim Q_{dw}: (\alpha_{dw2}, \beta_{dw2}, \gamma_{dw2}, t_{dw1}, t_{dw2}, t_{dw3}) \end{array}\right. \tag{3}$$

In Formula (3), $(\alpha_{up1}, \beta_{up1}, \gamma_{up1}, t_{up1}, t_{up2}, t_{up3})$ represents the pose of the upper camera, and $(\alpha_{dw2}, \beta_{dw2}, \gamma_{dw2}, t_{dw1}, t_{dw2}, t_{dw3})$ represents the pose of the lower camera. Since the stereo vision structure is fixed, there is a deterministic pose relationship between the two cameras. This pose relationship can be calculated using the stereo vision spatial control point mapping relationship, expressed as:

$$\left[\begin{array}{l} R_c = R_r R_l^T \\ T_c = T_r - R_c T_l \end{array}\right. \tag{4}$$

In Formula (4), R_c is the rotation matrix between the two cameras, and T_c is the translation vector between the two cameras, which can be calibrated separately using Zhang's calibration method. Therefore, by solving the camera pose above, the pose between the world coordinate system and the camera coordinate system can be obtained, and the local positioning information of the spatial target point can be calculated.

3.3 Global Association Localization of Out-of-View Reference Points

Establishing a rigid body transformation relationship between two coordinate systems through public point clouds is a key issue in establishing global associations between spatial points. Among the many solution methods, the SVD method directly obtains the global optimal solution through covariance matrix decomposition, which has obvious theoretical advantages. In contrast, the quaternion method is fast and avoids singularity, but requires additional processing of the translation component. The Kabsch algorithm, as a special case of SVD, is only applicable to rigid body transformation scenarios; while the ICP algorithm is insensitive to initial values and can handle partially overlapping point clouds, it risks getting stuck in local optima and has a high computational cost.

Based on the local region target point measurement and positioning, given the point set P{p_i} in the local coordinate system and the corresponding point set Q{q_i} (i=1,...,n) in the reference coordinate system, solve for the optimal rotation matrix *R* and translation vector *T* according to Formula (5) to minimise the objective function:

$$\min \sum_{i=1}^{n} \left\| (Rp_i + t) - q_i \right\|^2 \tag{5}$$

This paper uses SVD to solve *R* and *T*. Construct matrix *H* between local and reference systems:

$$\mathbf{H} = \sum_{i=1}^{n} M_i^{'} P_i^{'T} = M_1^{'} P_1^{'T} + M_2^{'} P_2^{'T} + \cdots + M_n^{'} P_n^{'T} \tag{6}$$

In Formula (6), $M_i^{'}$ ′ and $P_i^{'}$ represent the world coordinates centred on the centre of mass in the local coordinate system and reference coordinate system, respectively.

Perform singular value decomposition on *H*:

$$\mathbf{H} = \mathbf{U} \times \mathbf{S} \times \mathbf{V^T} \tag{7}$$

From Formula (6) and (7), after calculating the maximum value of $\sum_{i=1}^{n} y_i^T \mathbf{R} x_i$, we can obtain the optimal rotation matrix R as:

$$\mathbf{R} = \mathbf{VU^T} \tag{8}$$

Finally, substitute the value of *R* into formula (9) to calculate the translation vector *T*:

$$T = \bar{q} - R\bar{p} \tag{9}$$

The obtained R and T form the transformation matrix from local to global coordinates. Once a target point's coordinate is measured in the local system, it can be associated with the reference point using this transformation to compute their relative spatial position.

4 Experimental Verification

4.1 Experimental Apparatus Construction

By simulating the visual system layout parameters below, we can obtain the image factors affecting pose transformation accuracy. This allows us to derive principles for selecting the number of control points, focal length, and effective distance, which serve as the basis for system layout design.

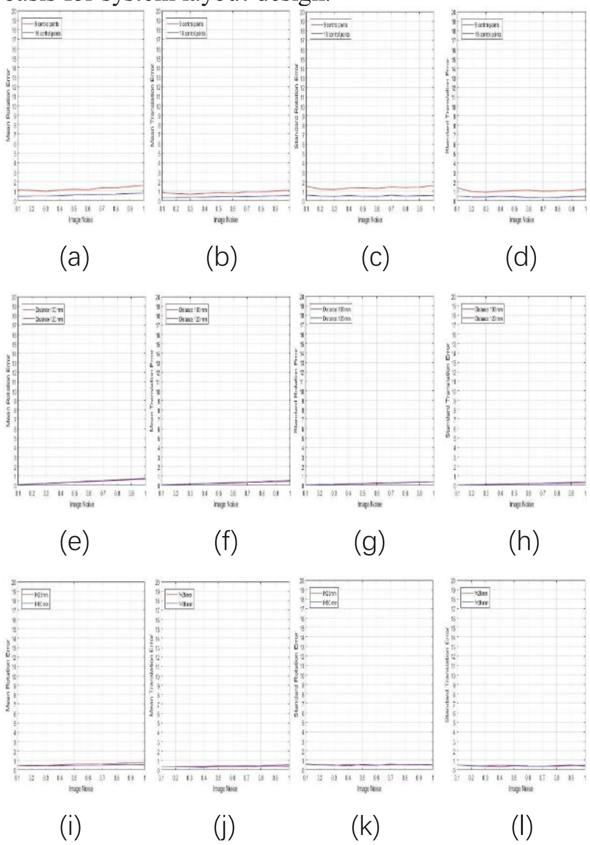

Fig. 3 Analysis of influencing factors of pose accuracy

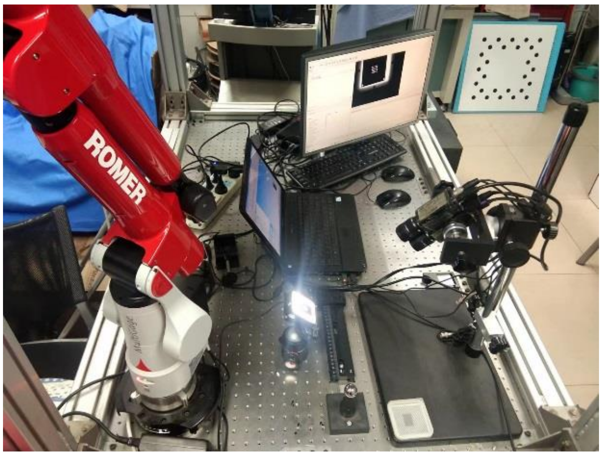

Fig.4. Experimental setup

Fig.5. Common point cloud construction

The experimental setup designed based on the simulation results is shown in Fig 4. The construction of the common control points is shown in Fig 5. The camera used is a Basler scA1600-14gm industrial camera, which provides the pixel coordinates of the centroid and the spatial three-dimensional coordinates of the target common points. The establishment of the reference points outside the field of view is performed on a ROMER articulated measuring machine. By measuring the control points and processing the data, their three-dimensional coordinates in the reference coordinate system are obtained.

4.2 Experimental Results

As shown in Fig 5, based on the three-dimensional coordinates of the target common points in the camera coordinate system and the three-dimensional coordinates in the reference coordinate system, the transformation matrix between the two coordinate systems can be calculated using these common coordinate point coordinate data. The calculation results are shown in Table 1 (displayed to 4 decimal places):

Table 1. Rotation Matrix

-0.9472	-0.1538	-0.2810
-0.2734	0.8451	0.4592
-0.1669	-0.5118	0.8426

Table 2 Cross-field Length Measurement Results (mm)

-93.4623
44.8897
-312.5442

Using this transformation matrix for cross-field geometric measurement (size > 300mm), results are in Table 3:

Table 3. Cross-field length measurement results (unit: mm)

No.	Measured Size	Actual Size	Relative Error
1	308.4134	308.3500	0.0634
2	311.3809	311.3778	0.0031
3	305.5944	305.6009	-0.0065
4	308.6417	308.5791	0.0626
5	311.5774	311.5354	0.0420
6	305.8562	305.8546	0.0016
7	308.7785	308.7487	0.0298
8	303.1281	303.1010	0.0271
9	306.0527	305.9858	0.0669
10	308.9711	308.8964	0.0746

As shown by the results, the relative error of the measurement method proposed in this paper is less than 0.1 mm. Preliminary experimental results indicate that even when the measured dimensions are completely outside the camera's field of view, the method proposed in this paper can still meet the requirements for high measurement accuracy.

5 Conclusion

To address the problem of global positioning of spatial points across fields of view, this paper investigates a local visual global positioning method that associates spatial points with out-of-field reference points. By analysing the characteristics of the target object, a parallel binocular visual local positioning model is constructed. using singular value decomposition to establish coordinate transformation relationships, thereby unifying the coordinate systems of local target points and global reference points. Preliminary results indicate that the method can achieve high measurement and positioning accuracy requirements, and the computational process involves matrix operations at the millisecond level, making it possible to perform rapid, precise, non-contact measurements of global geometric parameters for larger-sized workpieces.

Acknowledgments
The present study received financial support from the China Higher Education Society Project 23SZH0413, the National Higher Education Computer Basic Education Research Association Project 2024-AFCEC-460, and the Nankai University Educational Reform Project NKJG2025017.

References

1. R. Dabral, A. Mundhada, U. Kusupati, S. Afaque, A. Sharma, A. Jain, "Learning 3d human pose from structure and motion," in Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018, pp. 668-683.
2. X. F. Kong, Q. Chen, G. H. Gu, W. X. Qian, K. Ren, "GPS-based binocular camera calibration and space coordinate reconstruction," Opt.Precision Eng., vol. 25, pp. 485-492, Feb. 2017.
3. Y. C. Song, W. Chen, Y. T. Ye, J. X. Liu, B. Q. Huang, T. Guo, et al., "Sub-pixel stitching method for large object with duplicate texture of automatic visual inspection," Acta Optica Sinica, vol. 34, pp. 196-201, Mar. 2014.
4. D. H. Ji, D. Kim, J. Y. Kim, S. W. Byun, "Redundancy analysis and motion control using ZMP equation for underwater vehicle-manipulator systems," in IEEE Oceans, Shanghai China, 2016, pp. 1-6.

5. J. Santolaria, M. Ginés, "Uncertainty estimation in robot kinematic calibration," Robotics & Computer Integrated Manufacturing, vol. 29, pp. 370-384, Feb. 2013.
6. K. Ansari, P. Gyawali, P. M. Pradhan, K. D Park, "Coordinate transformation parameters in Nepal by using neural network and SVD methods," Journal of Geodetic Science, vol. 9, pp. 22-28, Jan. 2019.
7. O. Akyilmaz, "Total Least Squares Solution of Coordinate Transformation," Survey Review, vol. 1, pp. 68-80, Jan. 2007.
8. V. Lepetit, F. Moreno-Noguer, P. Fua, "Epnp: an accurate O(n) solution to the pnp problem," International Journal of Computer Vision, vol. 81, pp. 155-166, Feb. 2009.