An Improved SIFT Algorithm for Monocular Vision Positioning

Xinyue Niu a, *, Jiexin Pu b and Chi Zhang c

College of Information Engineering, Henan University of Science & Technology, Luoyang 471003, China.

*, a Corresponding author: nxy_sherl @163.com, bpujiexin@163.com, Czhangchi12306@163.com

Abstract. In view of the high real-time and accuracy requirements of the monocular hand-eye vision system in the positioning process, and in the case that the existing image matching algorithm can not meet these two requirements well at the same time, this paper improves the SIFT feature matching algorithm based on local features. First, the corner points determined in the Harris operator are used instead of the key points determined by the SIFT algorithm as feature points in the template image and the image to be matched. Then, a 32-dimensional feature description vector is constructed for each of the selected feature points through a Gaussian circular window. In the registration phase, the Euclidean distance is used as a measure function to match the 32-dimensional feature descriptors. Finally, 100 template images acquired by the monocular hand-eye vision experimental platform are used to test the matching effect of the improved SIFT algorithm, which proves that the improved algorithm has higher improvement in matching time and registration accuracy than the original algorithm. It is applicable to image registration for monocular vision positioning in industrial practice.

Key words: Image Matching; SIFT Algorithm; Monocular Vision; Harris Corner Points.

1. Introduction

In industrial production, vision is an important source of information for industrial robots to perceive the external environment[1]. The vision-guided positioning technology has good performances such as non-contact, high efficiency and fast dynamic response, which greatly improves the flexibility and operational flexibility of industrial robots[2-3]. Binocular stereo vision acquires the three-dimensional geometric information of the target point based on the principle of parallax, and the measurement accuracy of the target is high[4], but the hardware system is relatively complicated[5]. The monocular Eye-in-hand system has a simple structure, ensuring that the field of view of the industrial robot is unobstructed during the operation, and the detection area can be changed with the movement of the robot, It’s widely used in the grasping and placement, the peg-hole alignment and the high-precision self-assembly system[6-8].

Target detection is one of the core problems of monocular vision positioning[9]. Image matching is the key technology to achieve target detection. The matching algorithm directly affects the final effect of visual positioning[10]. The SIFT (Scale Invariant Feature Transform) algorithm based on local
features has the invariance of scale, rotation and translation, and can suppress the influence of noise and viewing angle change to a certain extent. It is a well-known image matching algorithm. However, the computational complexity of SIFT algorithm is high, and in most cases, the real-time requirements of industrial robots cannot be met. To this end, many scholars have invested in research to improve the SIFT algorithm. The literature uses principal component analysis (PCA) to represent high-dimensional data in low-dimensional subspaces and compress vector dimensions. This method effectively exploits the advantages of PCA technology, but increases the workload of training the projection matrix. In [15], the 60-dimensional square neighborhood descriptor is used to reduce the dimension of the descriptor, which increases the statistical range of the neighborhood pixel and enhances the uniqueness of the feature descriptor. The algorithm has improved in real-time, but it has high requirements for the environment of image acquisition, and it is not suitable for the case where the disturbance of working conditions in industrial production is uncertain. In [16], a circular window is used as a descriptor, and each feature point is represented by a 12-dimensional feature vector, which achieves a large dimensionality reduction, but the matching accuracy is not high in a complex scene. The literature considers that Harris corner points are simple to calculate and are not affected by illumination, rotation and noise. When improving the SIFT algorithm, the Harris algorithm is combined to achieve fast and robust image registration. However, when the unknown image rotation changes greatly, the mismatch rate increases.

In the case of fully considering the advantages and disadvantages of the SIFT algorithm, the corner points determined by the Harris operator are used to replace the key points determined by the SIFT algorithm in the template image and the image to be matched, so that these feature points are equipped with Harris anti-radiative transformation and the rotation invariance of SIFT at the same time. Then construct a 32-dimensional feature descriptor for these corner points through a Gaussian circular window, and reduce the dimension of the feature description vector in the original SIFT algorithm. Finally, the RANSAC algorithm is used to eliminate the mismatch point. The experimental results show that the improved SIFT image matching algorithm can meet the requirements of practicality and real-time in monocular vision positioning.

2. SIFT algorithm
The SIFT algorithm remains invariant to scale scaling, rotation, and even perspective changes. It is the most stable algorithm among many feature-based image matching algorithms. The main flow of the algorithm is shown in Figure 1.

![Flow chart of SIFT algorithm](image)

**Figure 1.** Flow chart of SIFT algorithm

Feature point detection: construct a scale space, determine and detect extreme points, remove the unstable points, including low-contrast points and edge response points, and then the remaining extreme points are the determined feature points. Feature point description: This feature point is characterized by calculating the amplitude and direction of the feature point gradient, and three information of scale, position and direction are determined for all feature points. The corresponding feature descriptor is established for each feature point by using 128-dimensional vector. Feature point matching: Match the feature points extracted in the template image and the image to be matched with a similarity measure function, such as the Euclidean distance ratio.
3. Improved SIFT algorithm

In the process of feature point extraction, the SIFT algorithm has to convolute the image with the Gaussian kernel function multiple times, so there is a problem that the calculation amount is large and the time is long. Moreover, the matching precision of the algorithm is insufficient to some extent, and mismatch may happen during the process of feature point matching. Therefore, an improved SIFT algorithm is proposed in this paper. In the extraction of feature points, the key points determined by the SIFT algorithm are not used, instead, the corner points determined in the Harris algorithm are employed, and a 32-dimensional feature description vector is constructed for these corner points through a Gaussian circular window to reduce the dimension of the feature description vector in the original SIFT algorithm. At the same time, the RANSAC algorithm is used to remove the mismatching point, which improves the matching precision of the algorithm.

3.1. Basic Steps of the Improved SIFT Algorithm

(1) The Harris operator is used to extract the corner points of the template image and the image to be matched, respectively. A set of corner points are established.

(2) For all the corner points determined in (1), the Gaussian circular window is used to construct a 32-dimensional feature descriptor for each of them.

(3) The Euclidean distance ratio is employed to match the feature descriptor determined in both the template image and the image to be matched.

(4) The RANSAC\(^\text{[21]}\) algorithm is used to remove the mismatched points.

3.2. Harris Operator for Corner Extraction

The corner extraction of the Harris operator is determined by equations (1) and (2).

\[
R = \det M - k(\text{trace}M)^2
\]

\[
M = \begin{bmatrix}
I_x^2 & I_x I_y \\
I_x I_y & I_y^2
\end{bmatrix}
\]

Where \(I_x\) is the derivative of the point \((x, y)\) in the \(x\) direction in the image; \(I_y\) is the derivative of the point \((x, y)\) in the \(y\) direction in the image, \(a\) and \(b\) are the two eigenvalues of \(M\). In the formula (1), \(\det M\) represents the determinant of \(M\), which is equal to the sum of \(a\) and \(b\). \(\text{trace}M\) represents the trace of \(M\), which is equal to the product of \(a\) and \(b\). \(k\) is a constant (typically 0.04-0.06). When the value of \(R\) is greater than a certain threshold value and a local extremum is obtained within a certain neighborhood, it is marked as a corner point.

3.3. Establishment of a 32-Dimensional Feature Descriptor

In this paper, a Gaussian circular window is used to establish a 32-dimensional feature description vector for selected feature points, which reduces the dimension of feature descriptors in the original SIFT algorithm.

Figure 2 is a schematic diagram of establishing the neighborhood of feature points. In the algorithm, the feature point is used as the origin, \(\theta\) is the polar angle to construct a two-dimensional coordinate system, which is divided into 32 sub-regions by the feature point neighborhood, and the feature vectors \(TH(x, y) = (HR_1, HR_2, \ldots, HR_{32})\) are standardized, the 32-dimensional feature descriptor is created for each feature point. Moreover, after being normalized, the feature descriptor has good adaptability to the unknown image that has been fuzzy transformed.
\[ HR_i = \frac{1}{NR_i} \sum_{(x,y) \in R_i} \text{trace}(x,y) \]  

(3)

Where \( NR_i \) represents the sum of the pixel points contained in each sub-area, and \((x, y)\) is the two-dimensional coordinates of the pixel points.

\[ \theta = a \tan \left( \frac{y - y_c}{x - x_c} \right) \]  

(4)

Where \( x_c \) and \( y_c \) are the coordinates of the feature points.

4. Experimental results and analysis

The experimental environment is Intel (R) Core (TM) i5 -2450M CPU@2.5GHZ processor, 4.00GB memory, simulation platform Matlab2010b, operating system Windows7. 100 template images were collected from the experimental platform of monocular hand-eye vision, including rotation transformation and scaling transformation fuzzy transformation. The collected template images were used for simulation, the performance of the proposed algorithm is evaluated in terms of total running time of the algorithm and matching accuracy.
The 100 collected template images were used for experiments, and the experimental results reveals that the improved algorithm is about 20% to 40% of the original SIFT algorithm in matching time (see Figure 4), and the registration accuracy rate (see Figure 5) is significantly higher than the original algorithm.

5. Conclusion
The improved SIFT algorithm uses the corner points determined by the Harris operator as feature points, avoids the large amount of convolution calculations required by the original SIFT algorithm, reduces the computational complexity, At the same time, the 32-dimensional feature descriptor is constructed by Gaussian circular window, which reduces the dimension of the original SIFT algorithm descriptor. Experiments show that the time consumed by the improved SIFT algorithm matching is 20% to 40% of
the original algorithm, and the matching accuracy is significantly improved compared with the original algorithm.

Acknowledgments
This paper was completed under the cordial care and careful guidance of Professor Pujiexin. His serious scientific attitude, rigorous academic spirit, and work style of excellence have deeply infected and inspired me. I would like to extend my sincere gratitude and lofty respect to him.

At the same time, I would also like to thank my classmates for their support and help in the details of the paper writing and modification process.

References
[1] M. Ito, M. Shibata. Visual tracking of a Hand–Eye robot for a moving target object with multiple feature Points: translational motion compensation approach[J]. Advanced Robotics, 2011, 25(3-4): 355–369.
[2] Ito, M., & Shibata, M. Visual servo control admitting joint range of motion maximally[C]. Lecture Notes in Control and Information Sciences, 2012:225–235.
[3] T. Kroger, J. Padial. Simple and robust visual servo control of robot arms using an on-line trajectory generator[C]. 2012 IEEE International Conference on Robotics and Automation, 2012:4862-4869.
[4] M. Kanbara, T. Okuma, H. Takemura, et al. A stereoscopic video see-through augmented reality System based on real-time vision-based registration[C]. IEEE Virtual Reality 2000, 2000:255–262.
[5] E. Royer, M. Lhuillier, M. Dhome, et al. Monocular vision for mobile robot localization and autonomous navigation[J]. International Journal of Computer Vision, 2007:74(3):237–260.
[6] C.-L. Shih, Y. Lee. A simple robotic eye-in-hand camera positioning and alignment control method based on parallelogram features[J]. Robotics, 2008, 7(2):31-37.
[7] S. Huang, Y. Yamakawa, T. Senoo, et al. Realizing peg-and-hole alignment with one eye-in-hand high-speed camera[C]. In Proceedings of the IEEE International Conference on Advanced Intelligent Mechatronics, 2013:1127–1232.
[8] W.-C. Chang. Robotic assembly of smartphone back shells with eye-in-hand visual servoing[J]. Robotics and Computer-Integrated Manufacturing, 2018, 50:102–113.
[9] T. Y. Wang, W. B. Dong, Z. Y. Wang. Position and orientation measurement system based on mmocular vision and fixed target[J]. Infrared and Laser Engineering, 2017, 46(4):1-8.
[10] E. Grosso, G. Metta, A. Oddera, et al. Robust visual servoing in 3-D reaching tasks[J]. IEEE Transactions on Robotics and Automation, 1996, 12(5):732–742.
[11] D. G. Lowe. Object recognition from local scale invariant features[C]. International Conference on computer vision, 1999:1150-1157.
[12] D. G. Lowe. Distinctive image features from scale-invariant key-points[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
[13] Y. Ke, R. Sukthankar. PCA-SIFT: a more distinctive representation for local image descriptors[C]. IEEE Computer Society, 2004:506-513.
[14] X. W. Wu, J. Cen, X. Y. Tai. SPCA: a fast dimensionality reduction method for image feature[J]. Journal of Ningbo University, 2005, 18(3):336-339.
[15] H. Liu, H. L. Shen. Image match method based on improved SIFT algorithm[J]. Micro Electrorns and Computer, 2014, 31(1):38-42.
[16] L. Liu, F. Y. Peng, K. Zhao, et al. Simplified SIFT algorithm for fast image matching[J]. Infrared and Laser Engineering, 2008, 37(1):181-184.
[17] L. Li, Q. L. Xie. An improved image matching algorithm based on Harris-SIFT[J]. Ship Electronic Engineering, 2017, 37(43):32-34.
[18] X. Q. Zhao, Y. F. Zhang. High-speed image matching algorithm matrix and improved based on trace of Harris autocorrelation gray scale value feature[J]. Journal of Lanzhou University of
Technology, 2018, 44(5):108-113.

[19] J. J. Xu, Y. Zhang, H. Zhang. Fast image registration algorithm based on improved Harris-SIFT descriptor[J]. Journal of Electronic Measurement and Instrumentation, 2015, 29(1):48-54.

[20] H. Zhao. Research on image registration algorithm based on point feature[D]. Shandong University, 2006.

[21] R. Raguram, J. M. Frahm, M. Pollefeys. A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus[C]. Computer Vision, 2008: 500–513.