Research on Augmented Reality Method Based on Improved ORB Algorithm

Ren Xiaokang¹, Cao Danling¹*, Ren Jie² and Du Bianli¹

¹ College of Computer Science and Engineering, Northwest Normal University, Lanzhou, Gansu, 730070, China
² School of Electronic and Information Engineering, Lanzhou City University, Lanzhou, Gansu, 730070, China
*Corresponding author’s e-mail: cdl_lll@163.com

Abstract. Aiming at the problems of time-consuming three-dimensional registration algorithm in augmented reality system, low accuracy of feature point matching and poor registration effect, an improved ORB (S-ORB) algorithm is proposed and applied to the Majiayao painted pottery augmented reality system. Firstly, the scale space is established by Gaussian filtering and downsampling, and the feature points are extracted by FAST corner detection. Secondly, the descriptors of feature points are calculated by rBFEF algorithm. Finally, the feature matching set is obtained by hamming distance and the mismatching point is eliminated by RANSAC algorithm to complete the three-dimensional registration. Make the virtual information stack correctly added to the real scene. The experimental results show that the S-ORB algorithm is better than the ORB, SURF and BRISK algorithms in time and accuracy when the target position is rotated and scaled. It has better real-time and robustness in the AR system of the painted pottery.

1. Introduction
Augment Reality (AR) technology is a computer-generated virtual three-dimensional object, text superimposed into the real scene, thus realizing the enhancement of reality. It has three characteristics: virtual and real, real-time interaction and three-dimensional registration [1]. The 3D registration technology based on natural features can meet more application requirements. It is the mainstream technology in AR system [2]. Firstly, the camera is used to capture the real scene, and then the correlation algorithm is used to extract the feature points in the scene. Finally, according to the coordinate information. The conversion determines the position of the virtual information in the real scene, and realizes the virtual and real registration [3]. Researchers have done a lot of research on extracting feature descriptors for feature point matching, such as SIFT (Scale Invariant Feature Transform) algorithm [4], SURF (Speed Up Robust Features) algorithm [4], and BRIEF (Binary Robust Independent Elementary Features). Feature description algorithm [6], BRISK (Binary Robust Invariant Scalable Keypoints) algorithm [7], ORB (Oriented FAST and Rotated BRIEF) algorithm [8] and so on. The literature [9] proposed a sensor-assisted BRIEF descriptor to ensure the rotation invariance and improve the accuracy of feature matching. In [10], the SIFT-ORB-MRANSAC fusion algorithm is proposed to complete the feature point extraction matching and improve the computational efficiency of the augmented reality system. In [11], an adaptive scale-invariant feature matching method based
on data clustering is proposed to solve the problem of poor robustness of feature matching in SIFT algorithm.

In this paper, S-ORB (Scale-ORB) is used in the 3D registration process, image pyramid is established by Gaussian filtering and downsampling, and features are extracted by FAST corner detection [12]. The rotation-related BRIEF descriptor is used to describe the feature. The Hamming distance is matched with features and the RANSAC algorithm [13] is used to eliminate mismatched points, and the three-dimensional registration is completed to accurately superimpose the virtual information in the real scene.

2. AR three-dimensional registration method

The time consumption and matching accuracy of the algorithm in the augmented reality system affect the accuracy of virtual information presentation [14-15]. In this paper, the S-ORB algorithm is used to improve the problem of scale invariance of ORB algorithm. The whole process mainly includes three parts: feature point extraction, descriptor generation and feature point matching.

2.1. Image feature point extraction

A set of Gaussian filtered images with different scales is constructed, and the same-images of different scales are downsampled. Considering the scale invariance and the real-time of feature extraction, the algorithm uses three times Gaussian filtering and three times downsampling to establish sparse image pyramids[12]. Guaranteed real-time feature extraction and scale invariance, as shown in figure 1. The image is unchanged when Scale=0, one downsampling when Scale=1, and two downsamplings when Scale=2, thus constructing pyramids with different resolutions. After the image pyramid is established, the FAST feature points of each layer of image are calculated and accurately positioned, so that the extracted feature points not only have the position information of the image, but also the scale information.

![Figure 1. Building an image pyramid](image)

After the image pyramid is established, the FAST corner detection is performed. The principle is to determine the feature points by detecting the brightness of the center point and surrounding pixels. By giving a Bresenhan circle centered on the candidate point, each pixel on the circle is checked one by one. If there are \( N \) consecutive points that are lit (or darker) than the center, the center point is determined to be feature point as in equation (1).

\[
N = \sum_{x \in \text{circle}(p)} |I(x) - I(p)| < \varepsilon_d
\]  

(1)

Where \( I(x) \) is the pixel value of any point on the circumference, \( I(p) \) is the pixel value of the candidate point, and \( \varepsilon_d \) is the threshold value of the pixel value difference. The selection of the radius of the Bresenhan circle is a very important parameter. For the sake of simplicity and efficiency, the feature points of 9-16 are mainly used to select feature points, as shown in figure 2. The FAST
algorithm extracts many repeated feature points for different scale layers of the same image. Firstly, the feature points extracted by different scale layers of the image pyramid are restored to the corresponding original images, secondly inversely transform the feature point pixel coordinates according to the number of downsampling, and then interpolation processing, etc. find corresponding feature points in the original image, perform scale space localization to finally determine the stable point of the image [12].

\[ Q = \left( \frac{m_{00}}{m_{01}}, \frac{m_{10}}{m_{11}} \right) \]  

\[ m_{pq} = \sum_{x,y,r} x^p y^q I(x,y) \]  

\[ \theta = \arctan\left( \frac{m_{01}}{m_{00}} \right) = \arctan\left( \frac{m_{01}}{m_{10}} \right) \]  

2.2. Image feature point description

The description of the image feature points uses BRIEF operator to calculate the descriptors of the feature points. Select \( n \) pairs of points in a certain pattern around the feature points, compare the gray value of the \( n \) pair of points as the equation (5), and use the binary string as the descriptor of the feature point as the equation (6). Where \( p(x) \) represents the gray value of point \( x \), \( n \) select 256 pairs of point with low correlation for each dimension and good discrimination. Resolving the rotation invariance by using the rotation-related BRIEF descriptor (rBRIEF)

\[ \tau(p; x, y) := \begin{cases} 1: p(x) < p(y) \\ 0: p(x) \geq p(y) \end{cases} \]  

\[ f_n(p) := \sum_{i=0}^{255} 2^{-i} \tau(p; x_i, y_i) \]  

2.3. Image feature point matching

Since the image feature point descriptor is a binary string, a preliminary match is made using the Hamming distance as in equation (7). Where \( a_i \) and \( b_i \) are the values of the ORB descriptor at the \( i \) bit.

\[ \operatorname{Ham}(\text{ORB}_1, \text{ORB}_2) = \sum_{i=0}^{255} a_i \oplus b_i \]  

In order to improve the accuracy of the matching, the RANSAC algorithm [13] is used to filter out the mismatching pair to obtain a more accurate matching result. The RANSAC algorithm is to randomly sample the samples in the matching data set and the samples cannot be collinear, calculate the homography matrix, and then use this model to test all the data, and calculate the number of data points and projection errors that satisfy the model (ie, the cost). Function), if the model is the optimal
model, the corresponding cost function is the smallest. Otherwise, the method of continuous iteration is used to find the optimal parameter model. A point that does not conform to the optimal model is defined as an "outer point", that is, a point that is mismatched.

3. Experiment and result analysis

The experiment was based on Visual Studio 2017 and Opencv 3.3 using the C++ programming language for experiments. In order to verify the effectiveness of the proposed algorithm, a single image is selected for feature point matching experiments. The experimental results of S-ORB algorithm and SURF algorithm, BRISK algorithm and ORB algorithm are shown in figure 3. The experimental data is shown in table 1.

![Feature point matching result](image)

**Figure 3. Feature point matching result**

| Algorithm          | Target image feature point | Match image feature points | Total time (ms) | Match points |
|--------------------|---------------------------|---------------------------|-----------------|-------------|
| SURF               | 685                       | 465                       | 1186            | 24          |
| BRISK              | 380                       | 352                       | 499             | 97          |
| ORB                | 470                       | 472                       | 468             | 275         |
| S-ORB              | 470                       | 472                       | 564             | 174         |

It can be seen that the number of matching feature points of the SURF algorithm is small, the computational complexity is high, the time spent is more than twice that of the BRISK algorithm, and the number of matching points is small; the BRISK algorithm has lower computational complexity, faster calculation speed, and more matching points than the SURF algorithm; The ORB algorithm is faster than other algorithms, and the number of matching points is more than other algorithms. The S-ORB algorithm is faster than the SURF and BRISK algorithms, and the calculation speed is slightly slower than the ORB algorithm.

The PASCAL public dataset and the Gansu Provincial Museum collected datasets were used to extract and match the feature points, and compared with the SURF, BRISK, and ORB algorithms in terms of accuracy and time. The accuracy is matched by the image matching and the total Match the ratio of the pair of points. figure 4, figure 5 is the experimental results of the PASCAL data set, figure 6, figure 7 is the experimental results of the painted pottery data set.

![Feature point matching accuracy](image)

**Figure 4. Feature point matching accuracy**

![Average recognition time](image)

**Figure 5. Average recognition time**
It can be seen from figure 4 that the feature point matching accuracy of the S-ORB algorithm is higher than that of the ORB and BRISK algorithms. Compared with the SURF algorithm, the average matching accuracy of the first group of image data is lower, and the fifth and eighth groups are lower. The average matching accuracy of image data is basically equal, and the feature matching accuracy of the remaining sets of image data S-ORB is significantly higher than that of SURF algorithm. In figure 5, the BRISK algorithm starts from the fifth group of image data and has a long time-consuming recognition time. The average recognition time of the SURF algorithm is significantly lower than that of the BRISK algorithm, which is higher than the ORB and S-ORB algorithms. The ORB algorithm takes the least time. The average recognition time of the S-ORB algorithm is slower than that of the ORB algorithm. The average recognition time of the fifth group of image data is slightly higher than that of the SURF and BRISK algorithms, and the others are lower than the SURF and BRISK algorithms.

![Graph showing feature point matching accuracy](image)

![Graph showing average recognition time](image)

It can be seen from figure 6 that the accuracy of the S-ORB algorithm is higher than that of the SURF algorithm and the ORB algorithm. Compared with the BRISK algorithm, the average matching accuracy of the first set of painted pottery images is lower than that of the BRISK algorithm, and the average matching rate of the fifth set of painted pottery images. Equal to the BRISK algorithm, the rest are higher than the BRISK algorithm. In figure 7, the image recognition time of the S-ORB algorithm is significantly faster than the SURF algorithm and the BRISK algorithm. Compared with the ORB algorithm, the recognition time of the first three groups of painted pottery images is slower, and the rest is basically flat.

4. Application

Combined with the research method of this paper, the three-dimensional model of rendering pottery is drawn, and the video related to the pottery is downloaded. The different Majia kiln pottery images are identified, which can complete the tracking registration and accurately display the virtual information of the 3D model and video of the pottery. The experimental results are shown in figure 8.

![Augmented reality effect](image)

The experimental results show that the S-ORB algorithm is applied in the painted pottery AR system, which has good robustness and can present virtual information accurately and timely. It can meet the needs of the Majiayao painted pottery augmented reality system.
5. Conclusion
Aiming at the problems of time-consuming augmented reality 3D registration algorithm, low accuracy of feature point matching and poor registration effect, the S-ORB algorithm is faster than SURF and BRISK algorithms, and slightly slower than ORB algorithm. Better than the ORB algorithm, SURF algorithm and BRISK algorithm, the virtual information is accurately superimposed in the real scene.

Acknowledgments
Thanks to my teacher Ren Xiaokang, for my careful teaching during the learning process, and for the painstaking revision and improvement of the problems in the thesis. His serious scientific attitude, rigorous academic spirit, and work style of excellence have deeply infected and inspire me. I would also like to thank the lab classmates for their help and encouragement to help me solve my doubts and overcome difficulties until the paper is successfully completed.

References
[1] Azuma, R.T. (1997) A survey of augmented reality. J. Presence: Teleoperators & Virtual Environments, 6(4):355-385.
[2] Li, G., Gao, S.B., Pan, Z.G., et al. (2018) Research on augmented reality based on unmarked recognition. J. Journal of system simulation, v.30(07):191-197.
[3] Gong, Y., Seibel, E.J. (2016) Feature-based three-dimensional registration for repetitive geometry in machine vision. J. Journal of information technology & software engineering, 6(4).
[4] Lowe, D.G. (2004) Distinctive image features from scale-invariant keypoints. J. International journal of computer vision, 60(2): 91-110.
[5] Bay, H., Tuytelaars, T., Van Gool, L. (2006) Surf: Speeded up robust features. In: European conference on computer vision. Springer, Berlin, Heidelberg. pp.404-417.
[6] Calonder, M., Lepetit, V., Strecha, C., et al. (2010) Brief: Binary robust independent elementary features. In: European conference on computer vision. Springer, Berlin, Heidelberg. pp.778-792.
[7] Leutenegger, S., Chli, M., Siegwart, R.Y. (2011) BRISK: Binary robust invariant scalable keypoints. In: 2011 IEEE International Conference on computer vision (ICCV). pp.2548-2555.
[8] Rublee, E., Rabaud, V., Konolige, K., et al. (2011) ORB: An efficient alternative to SIFT or SURF. In: ICCV. pp.2.
[9] Yu, L., Ong, S. K., Nee, A.Y.C. (2016) A tracking solution for mobile augmented reality based on sensor-aided marker-less tracking and panoramic mapping. J. Multimedia Tools and Applications, 75(6):3199-3220.
[10] Li, Y., Feng, N., Tan, S.C. (2019) Augmented reality method based on KCF and ORB improvement. J. Computer engineering, 45(8).
[11] Zhang, F., Gao, Y., Xu, L. (2019) An adaptive image feature matching method using mixed Vocabulary-KD tree. J. Multimedia Tools and Applications, 1-19.
[12] Gui, Z.W., Chen, J., Liu, Y., et al. (2014) A real-time scene recognition algorithm for smartphones. J. Journal of automation, 40(1):83-91.
[13] Fischler, M.A., Bolles, R.C. (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. J. Communications of the ACM, 24(6): 381-395.
[14] Shi, Q., Wang, Y.T., Chen, J. (2018) An augmented reality 3d registration algorithm based on visual sink. J. Chinese journal of image and graphics, 7(7):679-683.
[15] Wei, H.S., Huang, W.J., Dong, Q., Liu, Y.L. (2019) Shadow detection of outdoor video from the mobile perspective of augmented reality. J. Journal of computer aided design and graphics, 31(06):997-1006.