Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Robust Identification of Visual Markers Under Boundary Occlusion Condition*

Ruijie Chang, Yanjie Li, Chongying Wu
School of Mechanical Engineering and Automation
Harbin Institute of Technology, Shenzhen
Shenzhen, Guangdong Province, China
ysu_changrj@163.com , autolyj@hit.edu.cn,
15770627361@163.com

***Abstract* – Visual markers are widely used in indoor marker-based positioning systems to achieve higher speed and accurate positioning performance. However, the classic mark identification methods have certain limitations when encountering complex conditions. Especially, if the marker's boundary is blocked, it is almost impossible to identify the marker. In this paper, we redefine the identification task to a classification task based on CNN method. We also do some image transformations to create the dataset for our task. We train our dataset by transfer learning based on Google's Inception-V3 CNN model. The experimental results show that the classification method can handle the boundary occlusion problem well, which is also proved to be useful for other complex conditions.**

Index Terms – occlusion, visual marker, CNN, image processing.

I. INTRODUCTION

Visual markers are playing an important role in visual positioning system. Many intelligent systems achieve precise positioning with the help of visual markers. In augmented reality (AR) systems [1], we need use visual markers to track the objects in scene. In assistive robotics applications, visual markers make it easy to achieve camera calibration. In SLAM field [2], it's a popular way to achieve much faster and more accurate positioning with the help of visual markers.

In outdoor scenarios, making use of global positioning system (GPS) is the standard for positioning and navigation. However, in indoor GPS denied environments, it's also necessary to get the position and navigation information, the most popular ways are the computer-vision-based approaches. The computer-vision-based approaches need to deduce the position and the orientation of the camera. In computer-vision-based approaches, there are two main ways: marker-less approaches [3], [4] and marker-based approaches [5], [6], [7]. Marker-less approaches have more intelligent than marker-based approaches with no interference to the environment. Marker-based approaches are usually used when markers are undesirable because of aesthetic reasons. They need to get some pre-knowledge about the environment and also need to create some database about the features of the images caught by the camera, so that we can get the position and location of the camera. However, since they need some pre-knowledge about the environment, the marker-less systems are really computationally intensive and will consume a lot of CPU resources. Moreover, they fail in featureless places (such as corridors) and in environments with many repeated scenes (such as office). Until now, marker-less approaches can't satisfy many scenes with high speed and high accuracy. Therefore, in order to solve the problems faced by the marker-less approaches, the marker-based systems have been introduced, especially in the case of high speed and positioning accuracy requirements (such as smart storage systems). It is gratifying that the marker-based approaches perform very well.

The most popular kind of 2D marker is QR code. Ecklbauer (2014) showed that under different light conditions and for different size of markers, the detection rate of QR code is not suitable for speed required scenes [8]. Now, the mainstream square visual markers consist of black borders and internal code information. The four corners of these square visual markers are enough for camera calibration and pose estimation. However, in many practical applications (such as smart storage systems), when there exists occlusion, especially the occlusion of the border, many systems fail to recognize the visual markers. Artur Sagitov (2017) compared ARTag, AprilTag and CALTag in presence of occlusion [9], classic markers such as AprilTag have poor performance in this situation. Although CALTag is better in occlusion situation, it needs to add too much redundancy, which makes them rarely used in practical applications. A feasible idea is to improve the detection algorithm, which can cope with this kind of scenario. Víctor Mondejar-Guerra (2018) proposed a novel approach to handle the identification of markers, and transformed the identification problem as a classification one [10]. This kind of approach performs well under a wider range of environments. However, this approach only tests various single scenes one-by-one. The actual environment is often a combination of multiple complex scenes, so this method still needs some improvement.

The main contributions of this paper are as follows:

- Using the actual experiment, we prove that we don't need to add any redundant for the markers if we only

* This work is supported by Shenzhen basic research program JCYJ20180507183837726, JCYJ20170811155028832 and National Natural Science Foundation (NNSF) of China under Grant U1813206 and 61977019.

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

want to solve the boundary occlusion problem. Unlike CALTag, just do some change for the identification method, we can also achieve a good result to solve the boundary occlusion problem.

- A modified CNN-based classification method is used to identify the visual markers, which greatly improves the robustness of the visual markers in the boundary occlusion environment.
- Based on previous work, we add random boundary occlusion in image processing for the generation of training set. We generate an improved data set for deep learning of the visual markers. The experiment proves that this dataset can better cope with the boundary occlusion of visual markers.

II. RELATED WORK

In this section, we discuss some previous works about visual markers and the use of CNN method.

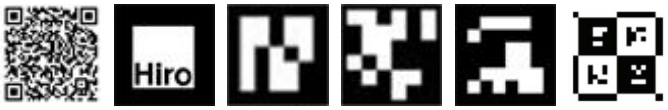
A. Visual marker systems

The visual marker system is composed of many visual markers, which confirms a dictionary. Many computer vision algorithms are used for recognizing and identifying the markers in the dictionary. Fig. 1 shows some of common visual markers in literatures.

The most common and widely used visual marker is QR code [11], which can be found in almost every corner of our life. It was invented in 1994 by Denso-Wave Corporation of Japan. Compared with other markers, the QR code has the advantages of fast reading speed, high data density and small footprint. The Reed-Solomon error correction algorithm makes it is possible for accurate identification even when more than 30% of markers' information has been lost. However, the identification of QR code has some requirements for camera resolution. Moreover, the QR code can't realize real-time identification, which makes it not a good choice for localization task [6].

Generally, visual markers achieve real-time effects by reducing the amount of information stored, and a binary code is enough. ARToolKit marker (Kato and Billinghurst, 1999) is one of the earliest visual marker systems [12]. It is typically surrounded by a black border with the custom pattern inside. The ARToolKit marker is widely used in Augmented Reality (AR) applications and can provide fast and accurate mark tracking. However, some shortcomings limit its application scenario. Firstly, ARToolKit marker is very sensitive to lighting conditions because of the fixed global threshold detection method. Secondly, as the number of markers increases, the identification accuracy declines dramatically and the computation cost increases simultaneously. Lastly, unlike other codes, the ARToolKit marker isn't encoded in binary but some symbols (for example, directly use Latin characters like "A"), which makes it's difficult to generate templates that related to each other.

ARTag (Fiala, 2010) is another widely used kind of visual marker, and employs binary codes for identification [13]. With some redundant bits for error detection, this kind of marker is robust with light change and some other conditions. Rather than the global threshold, ARTag adapt a new square detection method based on edge segmentation. The drawback of this kind of marker is its dictionary is fixed. Although it has considered redundant bits, the distance of each markers in the dictionary is not optimal, the markers are still very easy to confuse [10].



(a) (b) (c) (d) (e) (f)

Fig. 1 Common visual markers: (a) QR-code; (b) ArtoolKit; (c) ARTag; (d) AprilTag; (e) ArUco; (f) CALTag

Many other markers, such as ARToolKitPlus (Wagner & Schmalstieg, 2007), it adds a heuristic method to update global threshold to handle the single light problem [14]. Although ARToolKitPlus is a later work, it is not suitable in many other scenarios and the encoding system is worse than ARTag [6]. BinARyID (Flohr & Fischer, 2007) based on heuristic binary codification generates marker dictionaries [15], and can well deal with rotation ambiguity. However, BinARyID can't guarantee the robustness and the number of markers is limited [10].

AprilTag (Edwin Olson, 2011) is one of the most successful visual markers which is widely used in many occasions [6]. AprilTag system is composed by two parts: detector and the coding system. The special quad detector makes the system has a low false negative rate. The size of the dictionary varies with the size of markers, and makes the system has a good robustness. The distance of markers in the dictionary is large enough for a certain extent of error correction. One of the most obvious drawbacks of this system is the fixed dictionary, customers can't change dictionary with actual needs, the error correction ability still have space to optimize and the detection speed can be further improved [16].

ArUco (Garrido-Jurado et al, 2014) is another kind of most widely used visual marker [5]. Thanks to the kindness of the team, this system has a complete set of open source code from design to detection. Customers can design their own markers with actual needs. This system uses an adaptive thresholding method to increase robustness in complicated environment, and the error correction ability is not worse than AprilTag system. Their later work (Garrido-Jurado et al, 2016) generated an improved version ArUco marker system based on Mixed Integer Linear Programming method [16], which has the best error detection and correction abilities so far.

All the marker systems above have a common characteristic is that they all need to find the black border frame first to determine the position of the marker, so they are extremely sensitive to the occlusion of the boundary. Artur Sagitov (2017) had experimentally verified the sensitivity of the marking system to the boundary occlusion [9]. It is necessary to modify the detection algorithm so that it will have little or even no affection to the detection effect even if the border is occluded.

B. CNN-based classification

Nowadays, deep learning (DL) is leading an artificial intelligence revolution, can be seen in many research fields. Convolutional Neural Network (CNN) is a kind of DL, originally designed to solve image recognition problem. CNN is now used in many other tasks, such as video, audio signal and text data. CNN extracts spatial features using convolution. The basic components of modern CNN networks contains three parts: the convolutional layer, the pooling layer, and the fully connected layer. Four classic CNN architectures are very popular so far: AlexNet, VGGNet, Google Inception Net and ResNet, and their effect of classifying pictures has exceeded the human eyes.

Kazuya (2015) proposed a CNN-based finder pattern detection method [17]. Their work shows that even if the 2D codes is severely distorted or overlapped, they still can decode it. Although this article is aimed at QR codes, the method based on CNN offline learning is still worth learning. Víctor Mondejar-Guerra (2017) proposed the classification method to detect visual markers [10]. They verified that this method performs well in challenging conditions. However, their experiment is focused on many single conditions, our experiment shows the effect will be worse in a specific complex environment. What's more, too many environment pictures make the accuracy better than it should be, our work proves that the result is not so good if exists boundary occlusion.

III. Problem Induction

In this section, we introduce the typical methods of visual marker location task. Then we indicate the problem that this article needs to solve.

A. Typical methods

Typical marker-based location process uses image processing method. The process can be roughly divided into two steps: marker identification and camera calibration.

First, in marker identification process, it's necessary to confirm the existence of markers in the image, and the black external border plays an important role in this process. ArUco system uses a local adaptive method to determine thresholds, which can well cope with changes in illumination. AprilTag system uses a quad detection method. It makes use of the gradient amplitude of every pixel in the image to find the pixels that are on the same line segments. Then a recursive tree search method is adapted to find the line segments that are in the same quad. After confirm the existence of markers, it analyzes the internal area of the markers to obtain candidate marker information. Finally, discard those markers that are not in the dictionary and will realize identification. Notice that, if a candidate marker is not in the dictionary, but the distance with one marker in the dictionary less than the threshold, then think that the candidate mark is this mark and this process is called error correction. Different visual marker systems have different error correction abilities and this is an important indicator for the quality of a system.

The second step is camera calibration to calculate the angle and distance between the marker and the camera. We need to calculate a homography matrices of size 3×3, which projects the marker coordinate system into the 2D image coordinate system. Direct linear transformation (DLT) algorithms are commonly used to obtain homography matrices.

B. Problem introduction

This paper uses the classification method to deal with the visual marker recognition process under occlusion conditions. The most popular and accurate method nowadays adopts CNN to realize classification task. To achieve better classification results, a large number of data set is essential. The process of manually collecting data sets is too cumbersome and not very implementable. A practical method is to use image processing method to generate a lot of images. Then, we use these images to simulate the real situation that may encounters and integrate them into a set of datasets.

Select 50 markers from the ArUco dictionary as classification target. Combined with the complex conditions that often encountered in the actual scenes, we randomly perform one or more image transformation processes on the 50 markers. Finally, we also need to add a category to indicate that the captured images don't contain any markers. After the above analysis, our ultimate goal is a multi-class task in 51 categories.

IV. The Proposed Method

Based on the goals given in the previous section, we give the specific method of dataset creation in this section. Then we used a CNN-based identification method to handle the dataset.

A. Dataset generation process

As shown in Fig. 2, the complex situations that visual markers often faced in real-world scenes are defocus blur, motion blur, uniform light, overexposure, occlusion, etc. Denote that the size of each marker is $W\times H$, I(p) represents the intensity value of pixel p. Five processing methods are applied to the images: blur, shear, rotation, dilate and random occlusion.

(1) Image blurring is to smooth the image and smooth the sharper part of the image. Generally, this method can achieve image denoising. We can simulate defocus blur and motion blur if we select an appropriate blur kernel. The blur kernel equation is in equation (1).

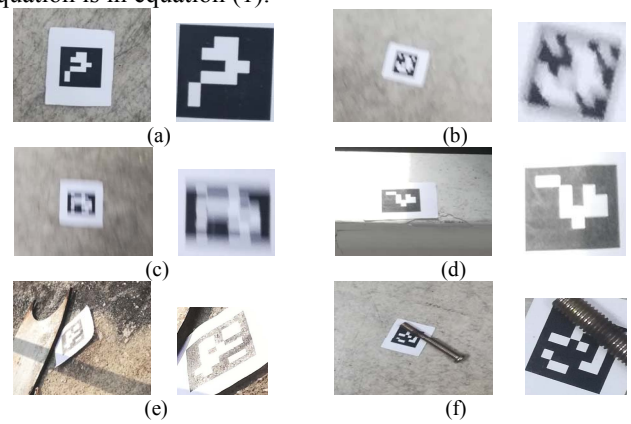


Fig. 2 Challenging conditions: (a) Ideal; (b) Defocus blur; (c) Motion blur; (d) Uniform light; (e) Overexposure; (f) Occlusion

(a) (b) (c) (d)

Fig.3 Relationship between blur effect and kernel size: (a) normal; (b) 1/10; (c) 2/10; (d) 3/10

$$K=\frac{1}{W\cdot H}\begin{bmatrix}1 & 1 & \cdots & 1\\ 1 & 1 & \cdots & 1\\ & & \cdots & \\ 1 & 1 & \cdots & 1\end{bmatrix}. \quad (1)$$

Where W is the image width, H is the image height. We use the method of mean filtering, also called box blur. The specific implementation process is to convolve the blur kernel and the original image, then we will get a blur image. As the convolution kernel increases, the blur effect gradually strengthens. The processing effect as shown in Fig. 3, the effect of the blurred kernels as 1/10, 2/10, and 3/10 of the image size. Two values W and H are randomly selected from $(1,\eta_1 s)$ as the two sides of the blur kernel. According to the final training effect, we take $\eta_1 = 0.15$.

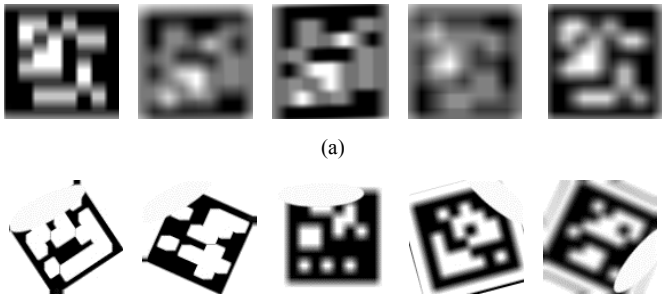
(2) Shear is a kind of affine transformation process. Shear contains horizontal and vertical shear, which can be used for deformation processing. Due to the shooting angle and other reasons, a situation we usually faced is that the photos we usually got will be deformed to a certain extent. Shear is a very good method to imitate this situation in real scenes.

The shear transform matrix are $\begin{bmatrix}1 & -\tan\alpha\\ 0 & 1\end{bmatrix}$ in x-axis direction and $\begin{bmatrix}1 & 0\\ -\tan\alpha & 1\end{bmatrix}$ in y-axis direction. The value of α following the uniform distribution $U(-\eta_2 s,\eta_2 s)$, where η_2=0.1 .

(3) Rotation is a situation that will inevitably exist. In the whole location process, we also need to determine and correct the camera position based on the rotation. In this article, we just randomly select a value from (0, 180).

(4) Image dilation is an operation that seeks a local maximum process. The structure element is convolved with the original image, and the maximum value of the coverage area of the structural element is calculated. The maximum value is assigned to the pixel that specified by the reference point, so that the image is highlighted gradually, achieving an effect similar to overexposure, the processing effect is shown in Fig. 4. The brightness increases as the size of the structural elements increases. In order to simulate different levels of exposure, the size of structural elements following the uniform distribution $U(1,\eta_3 s)$, where η_3=0.08 .

Fig. 4 Expanded image



(a)

(b)

Fig. 5 The generated images for training: (a) Training dataset image without edge occlusion; (b) Training dataset image with edge occlusion

(5) All the processing above are efficient for challenge conditions. However, when considering boundary occlusion problem, we also need some extra processing to improve the accuracy. Whether the visual marker code is occluded by other objects or the visual marker code itself is defaced, the visual marker code is occluded. In the deep training process, the random clipping method is often used to increase the training data set. It is also often good to achieve an effect similar to the occlusion of the image boundary. In this paper, the main starting point of using visual classification code recognition is that the visual mark code boundary is occluded, which makes it impossible to correctly identify the visual mark code. Therefore, in addition to the basic random cropping operation, we need to add additional operations to make the image occluded. Has the greatest impact on the results. The occlusion method used in this paper is to generate a fixed-size white ellipse, let it randomly occlude the visual marker code, and set its probability to be relatively large. This paper sets it to 0.9, so that 90% of the training data sets are obvious. The boundary occlusion condition makes it possible to finally achieve detection and recognition of the occlusion mark well.

Finally, the situations discussed above also occur according to a certain probability in the actual environment, and more may be the combination of multiple situations. According to others literatures and the research target of my own subject, the probability of occurrence of these cases are respectively set as: 0.75, 0.75, 0.75, 0.2, 0.9 respectively and Fig. 5 shows some generated markers for training.

B. A CNN-based identification method

In this paper, we transfer the typical identification task to a classification task. We use the CNN classification method to complete the identification goal. We use the transfer learning method to do the classification work. The transfer learning is a way to adjust the already trained model to be suitable for a new task. We use Google's Inception-v3 model to do our identification task. According to the conclusion in [18], we can realize classification task by replace only the fully connected layer of Inception-v3 model. We call this fully connected layer as bottleneck.

The first step in transfer learning is to adjust the number of basic dataset classes to the number of target dataset classes (51 classes in this article). The process of passing a new image through the trained convolutional neural network until the bottleneck layer can be regarded as a process of extracting

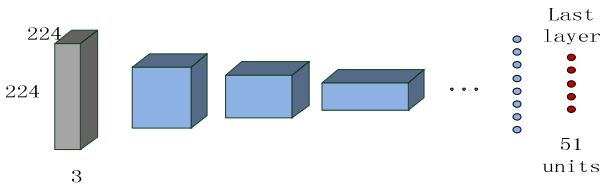


Fig. 6 Training process of our task

features from the image. Then, freezing and training, we just train the last layer to handle our identification task. The training process can be shown in Fig. 6. In Fig. 6, the original images' size are $224\times224\times3$, we do nothing to the previous layers. We directly use the trained Inception-v3 model to extract features from our dataset. Then use the extracted feature vector as input to train our new single-layer fully connected neural network to deal with the new identification problem.

V. Experimentation

This section introduces the training implementation process and demonstrates the effectiveness of our method. Comparing our work with previous articles, we verify the improvements of our article. At last, we analyzing the running time to illustrate the practical feasibility of our method.

A. Preliminary preparation

After the previous theoretical analysis, the analysis of the effect of using CNN classification method is carried out. We analyze the feasibility of this scheme from the recognition effect and recognition time respectively. For ease of analysis, we used 50 markers in the ArUco, and use each of which to generate 500 processed images respectively to form a training dataset with 25,000 images. Moreover, we use 50,000 images in MIRFLICKR-25000 as the environmental dataset. Finally, there are 75,000 images to make up the entire training data set, the picture is divided into 51 classification targets. In order to accurately analyze the actual recognition accuracy during training, another set of training that does not contain environmental variables is performed. At this time, the training data set has a total of 25000 pictures divided into 50 classification target. To verify the effect of our method for edge occlusion, we carry out two sets of experiment. Considering the real-time requirements in actual use, the trained network is analyzed for recognition time, compared with the original identification method, and the impact of this method on real-time performance is analyzed.

Our test data comes from the pictures taken in the actual scene, which can better illustrate the practical feasibility of the simulation data set generated by our simulation. In order to better reflect the effect of this method, the picture of the test data set is as fuzzy as possible. Unfavorable conditions such as distortion, and the number of pictures occluded by the boundary should be as many as possible, which can make the processing of the occlusion environment more convincing. Fig. 7 shows some test dataset images, and most of these test datasets contain labeled boundary occlusions, making the training results more targeted.

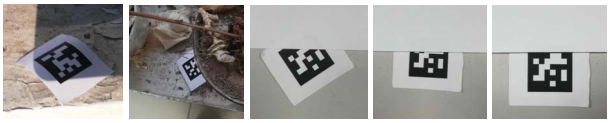


Fig. 7 Test examples in this paper

To ensure uniformity, we set all images size to 100×100 , and each test set is manually labeled to verify the accuracy of the test set. The markers is randomly shot and the size is not fixed. For convenience, we take 1000 images directly from the unused data of MIRFLICKR for testing.

B. Experimental result

In training process, we set the total training step is 10000, and 100 samples is randomly selected to test the accuracy of validation set every 100 steps. Firstly, the effect of adding environment categories on the experimental results is analyzed. The results are shown in Fig. 8. From the Fig. 8, we can see that since the environmental variables are very different from other categories, the classification is simple, so the overall accuracy will be made to be high, and it's difficult to distinguish the effect of our method for boundary occlusion. In order to accurately analyze the effect of CNN classification algorithm for marker recognition, we continue to design a training process that does not contain environmental variables.

We use the training dataset with occlusion and the training dataset without boundary occlusion, which can more directly reflect the important effect of adding occlusion processing on the recognition of boundary occlusion markers. The training results are shown in Fig. 9. As can be seen from Fig. 9, the accuracy can be obtained without adding boundary and occlusion processing. As can be seen from the figure, the final accuracy can reach about 90%, indicating that the identification can achieve better robustness by using this classification method. From the results of adding boundary occlusion, it can be seen that for the test data set with much boundary occlusion, the training data set with boundary occlusion will be more targeted and achieve better test result,

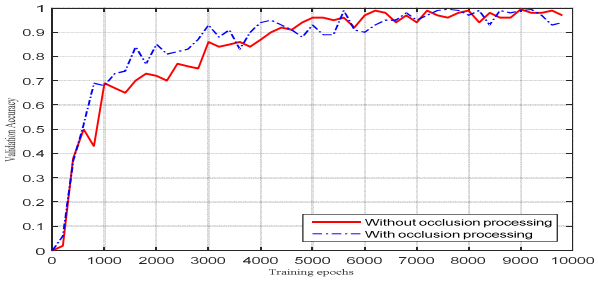


Fig.8 The training accuracy with environmental set

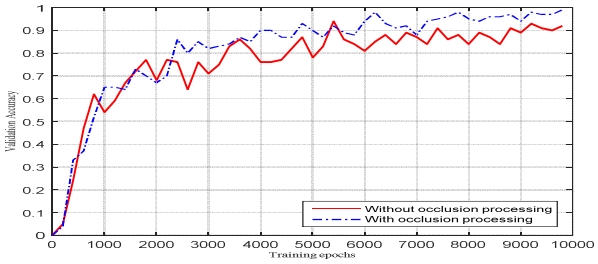


Fig.9 The training accuracy without environmental set

TABLE I
ACCURACY FOR TEST SET

Accuracy	With environment set	No environment set
The training set without boundary occlusion	99.2%	90.5%
The training set with boundary occlusion	99.5%	97.1%

TABLE II
COMPUTING TIME FOR EACH METHOD

	CNN	ArUco	AprilTag	CALTag
Time	6.15ms	0.096ms	4.52ms	\
Accuracy	97.1%	0	0	100%

the accuracy rate can reach about 97%. After all the training processes are organized, the test results are shown in Table I.

Finally, Table II lists the time of classification method and that of classic method to identify a single image. The result shows that the identification time of classification method is a little slower, but still can meet the real-time performance. It is not surprising that the accuracy of our method is not 100% as CALTag. We have not changed the form of the classic markers, but only improved the detection and identification method. If we add this method in classic method, then the positioning task will be more robust for our needs.

VI. CONCLUSION

This work has explored the identification process of visual markers under boundary occlusion conditions. Referring to the thoughts of previous article to transfer the identification task to a classification one, we have created our dataset for training process. Especially, the idea of adding random cropping during the creation process has greatly improved the identification effect under boundary occlusion condition.

This work also shows that in visual markers identification process, the size of environment dataset also has influence for the result. While the features of environment dataset are too obvious and the number of this dataset is much bigger than other classes', they will result in unreal high accuracy of experimental result. We compared our method of adding random cropping in training dataset generating process with previous method, the result shows that our method performs better than previous work.

Nonetheless, we also need to indicate that our method still has some limitations. First, we only verified the effect on recognition process, the positioning task also need the camera calibration process which we don't conduct in this paper. Second, although this paper improves the effect under boundary occlusion condition, the accuracy rate still has space to improve.

REFERENCES

[1] Chen P , Peng Z , Li D , et al. "An Improved Augmented Reality System Based on AndAR[J]." Journal of Visual Communication and Image Representation, 2015, 37:63-69.

[2] Verikas A , Radeva P , Nikolaev D P , et al. "SPIE Proceedings [SPIE Ninth International Conference on Machine Vision - Nice, France (Friday 18 November 2016)] Ninth International Conference on Machine Vision (ICMV 2016) - Comparative analysis of ROS-based monocular SLAM methods for indoor navigation[J]". 2017, 10341:103411K.

[3] S. Zhong, Y. Liu, and Q. Chen, "Visual orientation inhomogeneity based scale-invariant feature transform," Expert Systems with Applications, 2015, 42(13):5658-5667.

[4] Mur-Artal R , Montiel J M M , Tardos J D . "ORB-SLAM: a Versatile and Accurate Monocular SLAM System[J]." IEEE Transactions on Robotics, 2015, 31(5):1147-1163.

[5] Garrido-Jurado S , Mu?Oz-Salinas R , Madrid-Cuevas F J , et al. "Automatic generation and detection of highly reliable fiducial markers under occlusion[J]." Pattern Recognition, 2014, 47(6):2280-2292.

[6] Olson E . AprilTag: A robust and flexible visual fiducial system[C]// Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, 2011.

[7] Ababsa, Fakhr Eddine, and M. Mallem, "Robust camera pose estimation using 2d fiducials tracking for real-time augmented reality systems," Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry ACM, 2004.

[8] Ecklbauer, B. L., "A Mobile Positioning System for Android Based on Visual Markers," PhD Thesis, University of North Texas, Hagenberg, Austria, 2014.

[9] A. Sagitov, K. Shabalina, R. Lavrenov and E. Magid, "Comparing fiducial marker systems in the presence of occlusion," 2017 International Conference on Mechanical, System and Control Engineering (ICMSC), St. Petersburg, 2017, pp. 377-382.

[10] Mondéjar-Guerra, Víctor, et al. "Robust identification of fiducial markers in challenging conditions." Expert Systems with Applications (2017).

[11] Chu C H , Yang D N , Chen M S . "Image stablization for 2D barcode in handheld devices[C]"// Proceedings of the 15th International Conference on Multimedia 2007, Augsburg, Germany, September 24-29, 2007. ACM, 2007.

[12] H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99), San Francisco, CA, USA, 1999, pp. 85-94.

[13] M. Fiala, "Designing Highly Reliable Fiducial Markers," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 7, pp. 1317-1324, July 2010.

[14] Wagner, Daniel, and D. Schmalstieg, "ARToolKitPlus for Pose Tracking on Mobile Devices," Proc Computer Vision Winter Workshop 2007.

[15] Flohr, Daniel, and J. Fischer, "A Lightweight ID-Based Extension for Marker Tracking Systems," (2007).

[16] Garridojurado, S, et al. "Generation of fiducial marker dictionaries using Mixed Integer Linear Programming." Pattern Recognition, vol. 51, no. C, pp. 481-491, 2016.

[17] Nakamura, Kazuya, H. Kawasaki, and S. Ono. "Agent-based two-dimensional barcode decoding robust against non-uniform geometric distortion." The 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR), IEEE, 2015.

[18] J. Donahue, Y. Jia, O. Vinyals, et al. "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition," J. Computer Science, 2013.