OBJECT RECOGNITION ALGORITHM BASED ON IMPROVED K-MEANS BAG OF WORDS MODEL

Jianhui SONG[1], Xin SONG[2], Yanju LIU[3]*

An object recognition algorithm based on improved k-means bag of words model is proposed in this paper. The proposed algorithm extracts the SIFT feature from the target image. The k-means algorithm is applied to construct the bag of words model with SIFT feature point. The obtained bag of words model is fed into the support vector machine for training. Then the target recognition is realized. The proposed algorithm solves the shortcomings that the training of the high dimension and large number of SIFT feature points directly feeding into support vector machine takes too much time. The dimension of the image feature points is greatly reduced. The experimental result shows that the improved k-means bag of words model can better represent the bag of words model with SIFT feature point. The system training time of the proposed method is much shorter than the classic SIFT algorithm. The object recognition algorithm based on improved k-means bag of words model proposed in this paper has a higher recognition rate and stability in target recognition than the original k-means algorithm given the initial cluster centers.

Keywords: K-means, Bag of Words Model, Density Parameter, Object Recognition

1. Introduction

Target recognition plays an increasingly important role in contemporary society [1-3]. Feature extraction is the key technology of image object recognition. Commonly used image features are color features, texture features[4], shape features, and spatial relationships. According to the image features used in the recognition, the current object recognition method are mainly divided into two categories: method based on the key point of the image and method based on the object edge shape. The localization of the feature makes it possible to identify the obscured object or in a complex environment. Therefore, local feature-based target recognition becomes a hot spot in current visual research. Tinne Tuytelaars[5] made a review of local

[1] School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang, Liaoning, P.R. China. E-mail: hitsong@126.com.

[2] School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang, Liaoning, P.R. China

[3]* (Corresponding author) School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang, Liaoning, P.R. China

invariant descriptors over the last half century. Among them, the Scale Invariant Feature Transform (SIFT), which was first proposed by David G. Lowe in 1999 and completed in 2004, is most widely used in research[6]. As a large number of SIFT feature points can be extracted from each image and each feature point generates a 128-dimensional feature vector, the object recognition using the SIFT feature points has disadvantage of too large amount of data.

Li Feifei proposed a "bag of words" model (bow) [7], and made a very good performance in a classification task. It has the advantages of simplicity and ease of execution. However, its disadvantage is that the clustering results are not stable and the number of the clustering centers is not easy to determine. In order to improve the performance of bag of words model, the original k-means clustering algorithm is improved to construct the bag of words model.

k-means algorithm requires specific pre-given number of clusters k. Different k value will not only get a different bag of words model, but also affect the final target recognition results [8]. In the traditional k-means clustering algorithm, the initial cluster centers are selected by using multiple randomly method. However, using the multiple randomly method to choose the initial cluster centers faces with large amounts of data and cause a great waste of time [9-11]. The author of this article proposed an improved k-means algorithm [12].

In the case of target recognition classifier, discriminator model is the most intuitionist and important model, which mainly includes the following methods [13]: template matching method, statistical method, structure pattern recognition method, artificial neural network method and integrated learning method. As a more mature method in statistical learning and machine learning, support vector machine (SVM) takes the sample learning error as the constraint condition for optimizing the classification interface problem, and minimizes the confidence range value as the optimization target to be achieved.

The object recognition algorithm based on improved k-means bag of words model is proposed in this paper. The proposed algorithm extracts the SIFT feature from the target image. The k-means algorithm is applied to construct the bag of words model with the SIFT feature point. The obtained bag of words model is fed into the support vector machine for training. Then the target recognition is realized. The object recognition algorithm based on improved k-means bag of words model proposed in this paper is compared with the object recognition algorithm based on the original k-means bag of words model.

2. Target Recognition Algorithm Based on the Optimized K-means Algorithm

The SIFT feature point describes the local characteristics of the image. It has the characteristics of rotation invariance, scale invariance and brightness invariance. When the viewing angle changes, or there is noise, the SIFT feature point can maintain a certain degree of stability. The SIFT feature descriptor has a flexible extensibility, which can be easily combined with other forms of eigenvectors. Therefore, this paper extracts the SIFT feature point to identify the target. The target identification flow chart is shown in Fig. 1.

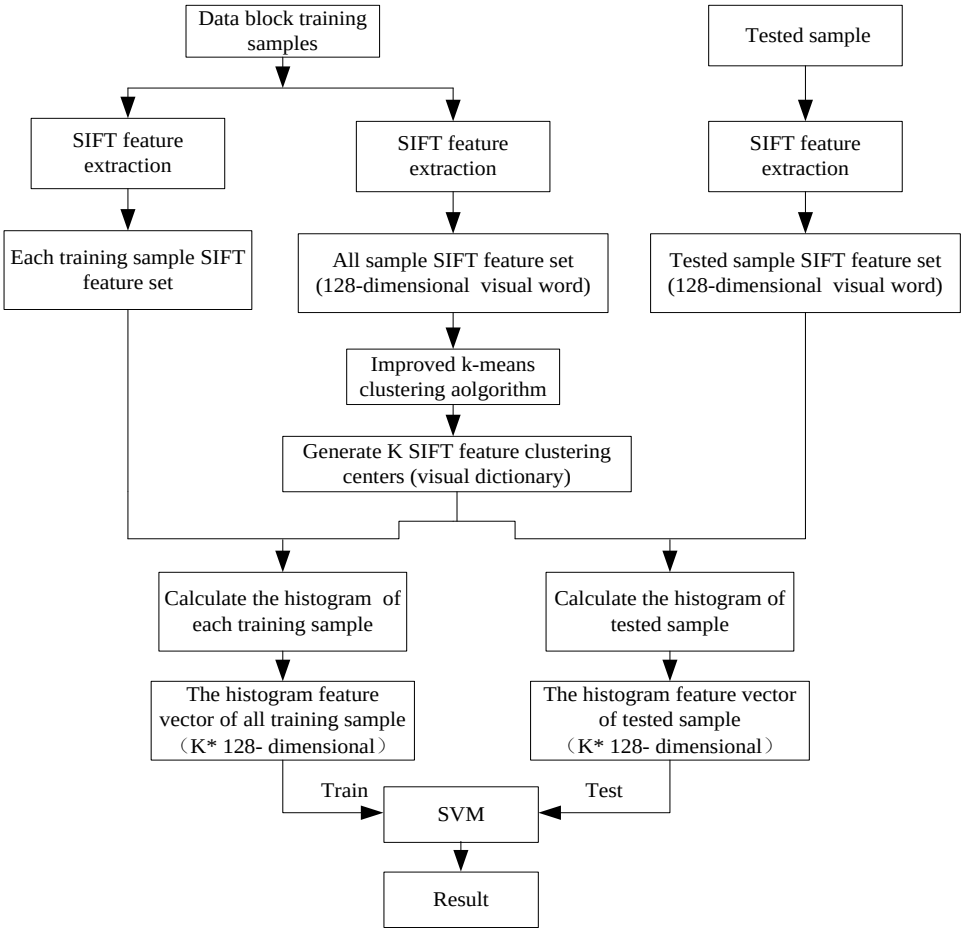


Fig. 1 Target recognition flow chart

The SIFT feature point feature vector is extracted from the training sample image. The improved k-means clustering algorithm is used to generate

the SIFT feature vectors clustering centers. Each cluster center is a 128-dimensional vector, which forms the visual dictionary. The distribution of the SIFT feature vector on each training sample in the visual dictionary is calculated. The histogram feature vector of each training sample is calculated. The obtained histogram feature vector of the training sample is fed into the support vector machine for training. The distribution of the SIFT feature vector on each tested sample in the visual dictionary is calculated. The histogram feature vector of each tested sample is calculated. The obtained histogram feature vector of the tested sample is fed into the support vector machine, and then the target recognition is realized.

2.1 The Establishment of Bow Model

The target recognition process consists of four parts: the image acquisition, image processing, feature extraction and feature selection [14]. The SIFT features are extracted to identify the target. This is a kind of local characteristic having very good scale invariance. However, each feature point has 128 dimensions. The feature points extracted from each class can be millions when feature extraction is performed on the selected images, which causes the classifier training behind inconvenience. Therefore, a good model is needed. In order to solve this problem, the bag of words model proposed by Li Feifei is adopted. The bag of words model is shown in Fig. 2. It can be seen from the picture that an image is composed of several pieces and each piece is a "word".

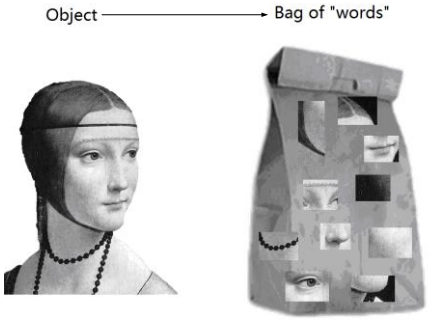


Fig. 2 "Bag of words" model

The construction process of bow model is shown in Fig. 3.

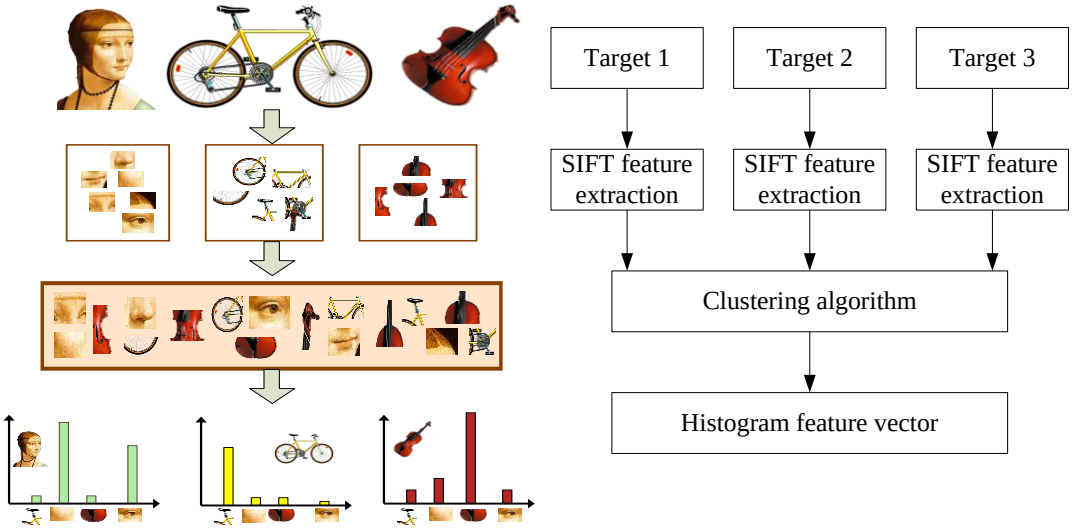
(1) A large number of SIFT feature points are extracted from each category model of the target model library as the "visual vocabulary", which is equivalent to the word in the document.

(2) The number of all the SIFT feature points are too large. If the SIFT feature points are directly used to describe the target template, the system

training will take too much time. Therefore, the SIFT feature vector extracted from all model images is clustered by k-means clustering algorithm. The clustering centers are used to form the visual dictionary.

(3) The SIFT feature points in each target class image is classified into the nearest clustering center. The distance metric adopts the Euclidean distance. The number of each word occurrences in the image is counted in the dictionary. Each target image is represented using a M-dimensional histogram feature vector.

In this paper, the improved k-means is used to build the bag of words model in target recognition and it is compared with the bag of words model that builds through original k-means algorithm.



(a) Algorithm diagram (b) Algorithm flowchart

Fig. 3 Construction process of bow model

2.2 The K-means Algorithm based on the Distribution of Maximum Density Points

The optimized k-means algorithm for selecting initial clustering center is shown in Fig. 4 [12]. The feature points having the same density have the possibility that the distance between them is greater than the average distance. Then the two points are used as candidate points of the initial clustering point. After k cycles, the resulting candidate points may be larger than the pre-set k value, then the first k points is selected as the initial clustering point through

the maximum distance method.

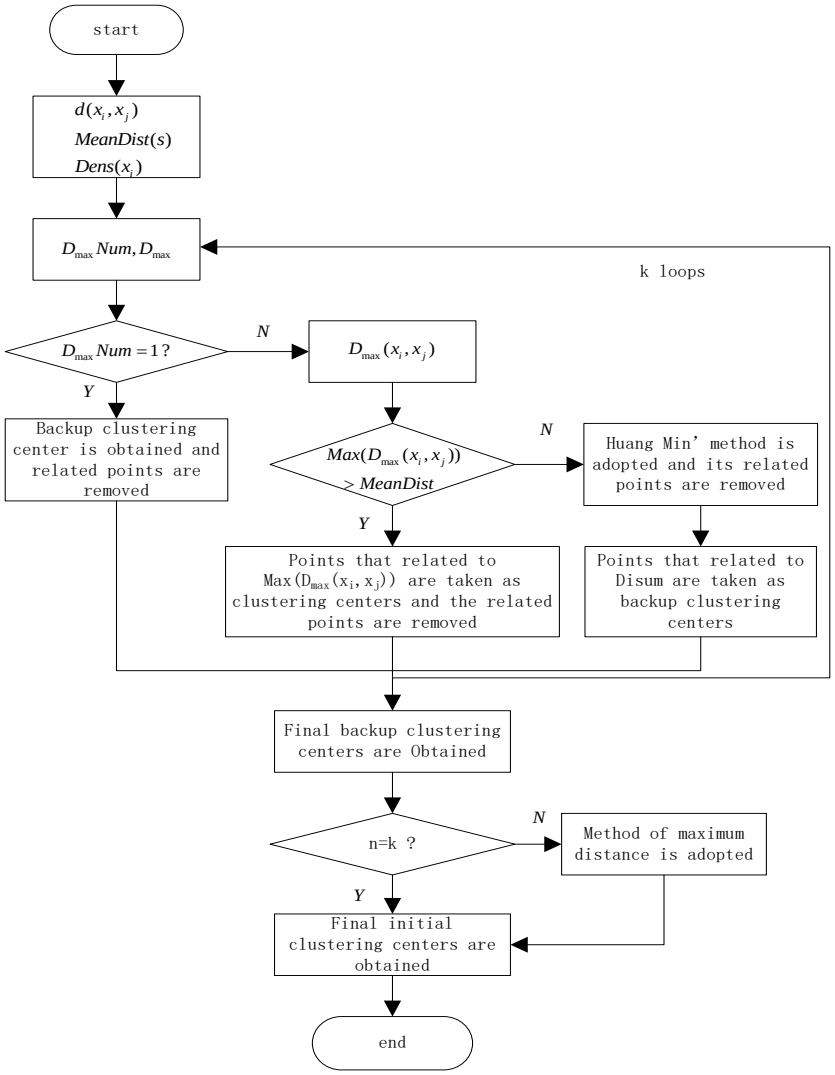


Fig. 4 Algorithm flow char

2.3 The Principle of SVM Classification

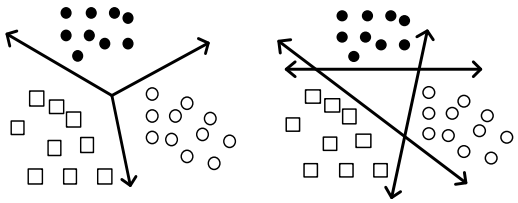
The libsvm software package developed by Taiwan University Professor Lin Zhiren is chosen in terms of classification. SVM is a kind of classifier which is widely used nowadays [15, 16]. When the sample is not linearly separable, it can be projected into a high dimensional space by a nonlinear mapping. Therefore, the high dimensional space can be linearly separable.

At present, there are two ways to construct multi-class support vector

machines using two-class SVM.

The first direct method is to modify the objective function directly and merge the parameters of multiple classification surfaces into one optimization problem. By solving this optimization problem, the multi-classification is solved at once. This method seems to be simple. However, its solution process is too complex and needs too much computation. It is also difficult to achieve. Therefore, it is not widely used.

The second one is the indirect method, that is, multiple binary SVMs are combined. There are two common methods: one-versus-one (OVO) and one-versus-rest (OVR). The simplified model diagram can be seen in Fig. 5.



(a) one-versus-one (b) one-versus-rest

Fig. 5 Models of OVO and OVR

(1) One-versus-one. For N-type training data set, every two N-type data will be pairs. Each pairs obtained a SVM model. And a total of $N(N-1)/2$ classifier is obtained. The advantage of this one-versus-one combination is that there is no need to modify the objective function. And there is no complex model and the solution process. The relation between number of samples and SVM is shown in Fig. 6. It can be seen that with the increase of N, the demand for the number of classifiers is larger. And the two are not linear.

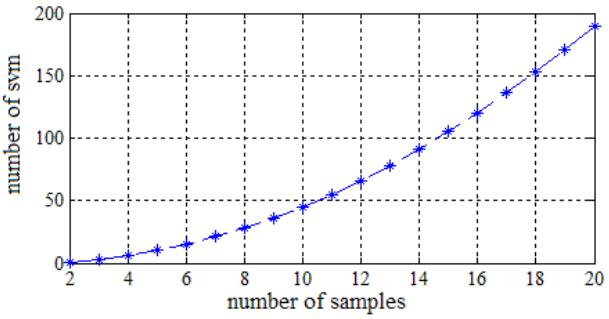


Fig. 6 Relation between number of samples and SVM

(2) One-versus rest. In the case of sample partitioning, the one-versus rest method need to classify a class of samples into one class, and the remaining samples into another class. Therefore, N-type training data set need N SVM model. The class with the largest classification function value is the target class. The advantage of this method is that the N class of samples only needs N SVM models. The method needs fewer models and the classification speed is relatively fast. However, if the clustering method is adopted to get the corresponding bag of words model, there will be a huge amount of data and the requirements of the equipment is relatively high.

According to the needs of the experiment, this paper will use one-versus-one classification method. The voting method is used to determine the results.

3. Experimental Results and Analysis

The ketch, minaret and windmill images used in the experiment are derived from the Caltech256. The number of images respectively is 80, 98 and 92. The extracted SITF feature point's number of ketch, minaret and windmill are shown in Table 1. Each feature point generates a 128-dimensional feature vector.

Table 1.

The extracted SITF feature points number

	ketch	minaret	windmill
Number of SITF feature points	34512	44660	52815

The number of all the SIFT feature points are too large. If the SIFT feature points are directly used to describe the target template, the target match will take too much time. Therefore, the SIFT feature vector extracted from all model images is clustered by k-means clustering algorithm. 200 clustering centers are generated from all of the SIFT feature points using the k-means clustering algorithm. Each cluster center is a 128-dimensional feature vector. The algorithm performance comparison based on SIFT feature points and the proposed method is shown in Table 2.

Table 2.

The algorithm performance comparison

	SIFT	SIFT+k-means bag of words
SVM training time(h)	19.6	8.5
Target recognition time(s)	0.37	0.26
Average recognition rate(%)	84	81

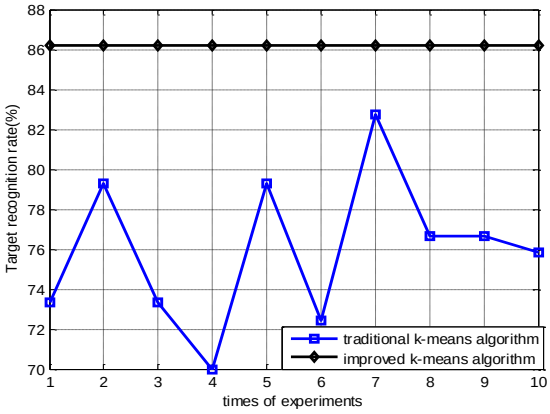
It can be seen from Table 2 that the target recognition rate of the proposed method is similar with the classic SIFT algorithm, but the system training time of the proposed method is much shorter than the classic SIFT algorithm. The clustering center obtained is taken as the visual vocabulary of the two categories and the corresponding bow model is obtained. A given number of training data is trained several times. In order to search for the best clustering centers, the initial clustering centers are randomly selected ten times for the groups of clusters. The ten groups are used as visual words and the corresponding bow model is applied to each image. Ten groups of training experiments are done. Each experiment is tested twenty times in training and testing. In the target recognition of ketch and minaret, the maximum, the minimum and the average recognition rate of the twenty test results are selected from each test set and written in Table 3. The results are compared with the object recognition algorithm based on improved k-means bag of words model proposed in this paper. The target recognition rate graphics for ketch and minaret using two algorithms are shown in Fig. 7. In order to facilitate comparison with the original algorithm for a given initial clustering center, the ten times is specially marked.

It can be seen from Table 3 and Fig. 7, in the target recognition of ketch and minaret, the maximum, minimum and average value of the classification results using the improved algorithm is better than the original algorithm given initial clustering center.

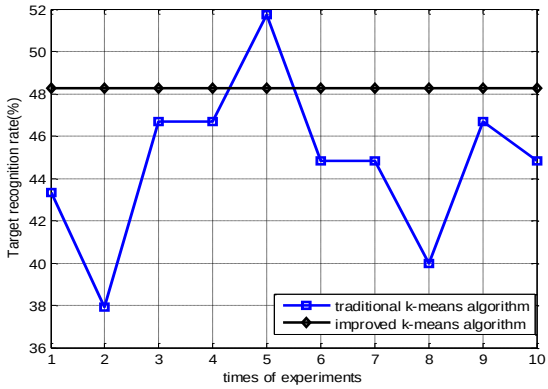
Table 3.

Target recognition rate for ketch and minaret using two algorithms

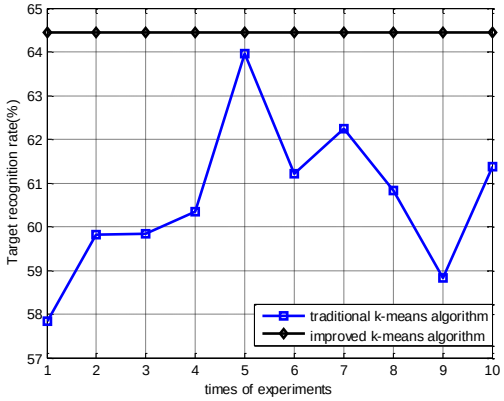
		Maximum (%)	Minimum (%)	Average value (%)
Primal algorithm (given initial clustering center number)	1	73.3333	43.3333	57.83333
	2	79.3103	37.931	59.82759
	3	73.3333	46.6667	59.83335
	4	70	46.6667	60.33334
	5	79.3103	51.7241	63.96551
	6	72.4138	44.8276	61.2069
	7	82.7586	44.8276	62.24137
	8	76.6667	40	60.83333
	9	76.6667	46.6667	58.83333
	10	75.8621	44.8276	61.37932
Improved algorithm		86.2069	48.2759	64.44275



(a) The maximum test results



(b) The minimum test results



(c) The average value test results

Fig.7 Target recognition rate graphics for ketch and minaret using two algorithms

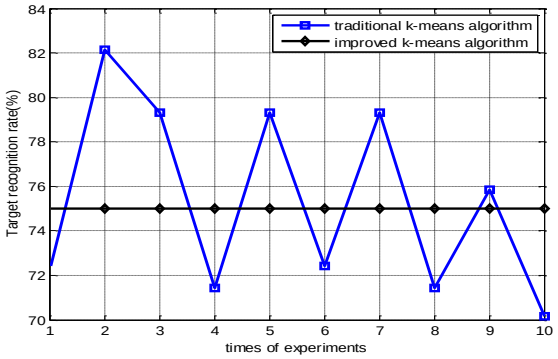
Although there is a point whose minimum is comparable to the improved algorithm. However, from the maximum and average value of the

corresponding point, it can be seen that the number of such points are few. The fluctuations of the original algorithm given the initial clustering center number are also relatively large. In the target recognition of ketch and windmill, the maximum, the minimum and the average recognition rate of the twenty test results are selected from each test set and written in Table 4. The results are compared with the object recognition algorithm based on improved k-means bag of words model proposed in this paper. The target recognition rate graphics for ketch and windmill using two algorithms are shown in Fig. 8.

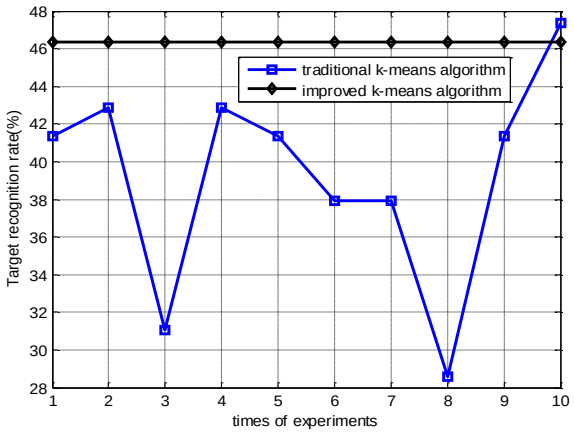
Table 4.

Target recognition rate for ketch and windmill using two algorithms

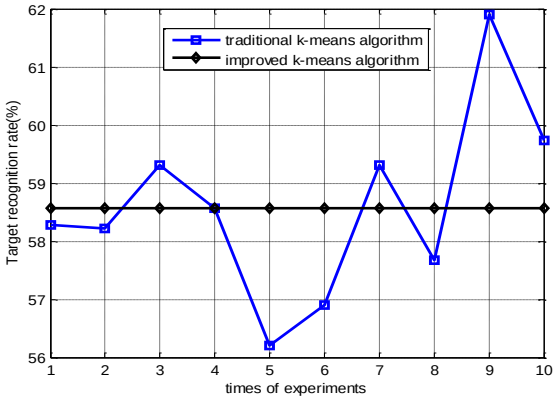
		Maximum (%)	Minimum (%)	Average value (%)
Primal algorithm (given initial clustering center number)	1	72.4138	41.3793	58.27586
	2	82.1429	42.8571	58.21429
	3	79.3103	31.0345	59.31034
	4	71.4286	42.8571	58.57143
	5	79.3103	41.3793	56.20689
	6	72.4138	37.931	56.89654
	7	79.3103	37.931	59.31034
	8	71.4286	28.5714	57.67858
	9	75.8621	41.3793	61.89656
	10	70.1754	47.3684	59.73686
Improved algorithm		75	46.3684	58.57144



(a) The maximum test results



(b) The minimum test results



(c) The average value test results

Fig. 8 Target recognition rate graphics for ketch and windmill using two algorithms

From Table 4 and Fig. 8, it can be seen that the classification results of the two algorithms are relatively close in terms of the maximum and average values. The effect of the improved algorithm proposed in this paper is better than the original algorithm given the initial cluster centers in the minimum value. In the target recognition of minaret and windmill, the results are similar to the target recognition of ketch and windmill.

4. Conclusions

The object recognition algorithm based on improved k-means bag of words model is proposed in this paper. In order to make better use of the SIFT feature to construct bag of words model, the improved k-means is used to build the bag of words model with the SIFT feature point in target

recognition. The proposed algorithm solves the shortcomings that the training of the high dimension and large number of SIFT feature points directly feeding into support vector machine takes too much time. The dimension of the image feature points is greatly reduced. The target recognition rate of the proposed method is similar with the classic SIFT algorithm, but the system training time of the proposed method is much shorter than the classic SIFT algorithm. The object recognition algorithm based on improved k-means bag of words model proposed in this paper is compared with the object recognition algorithm based on the original k-means bag of words model. Experimental results show that the object recognition algorithm based on improved k-means bag of words model proposed in this paper has a higher recognition rate and stability in target recognition than the original k-means algorithm given the initial cluster centers.

Acknowledgement

The study is funded by Shenyang Ligong University Key Laboratory Open Fund Project (Intelligent and Network Measurement and Control Technology Key Laboratory of Liaoning Province, Grant No.4771004kfs53).

REFERENCES

[1]. *Y. G. Sun, L. Du, Y. Wang, Y. H. Wang, and J. Hu,* "SAR Automatic Target Recognition Based on Dictionary Learning and Joint Dynamic Sparse Representation", IEEE Geoscience and Remote Sensing Letters, **vol**. 13, no. 12, pp. 1777-1781, 2016.

[2]. *Y. Y. Cheng, F. Hu, L. Q. Gui, J. L. Su, B. Qi, S. Y. Liu, and M. M. Huang,* "Linear polarisation property and fusion method for target recognition in passive millimetre-wave polarimetric imaging", Electronics Letters, **vol**. 52, no. 14, pp. 1221-1223, 2016.

[3]. *Z. T. Yang, J. Deng, and A. Nallanathan,* "Moving Target Recognition Based on Transfer Learning and Three-Dimensional Over-Complete Dictionary", IEEE Sensors Journal, **vol**. 16, no. 14, pp. 5671-5678, 2016.

[4]. *B. G. Batchelor, F. M. Waltz, "Intelligent machine vision: techniques, implementations and appliations", Springer-Verlag, London, pp.31-64, 2001.*

[5]. *T. Tuytelaars and K. Mikolajczyk,* "Local Invariant Feature Detectors: A Survey", Foundations and Trends in Computer Graphics and Vision, **vol**. 3, no. 3, pp. 177-280, 2008.

[6]. *D. G. Lowe,* "Object recognition from local scale-invariant features", Proceedings of the Seventh IEEE International Conference on Computer Vision, **vol**. 2, pp. 1150-1157, (1999)

[7]. *F. F. Li and P. Perona,* "A bayesian hierarchical model for learning natural scene categories", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **vol**. 2, no.2, pp. 524-531, 2005.

[8]. *S. B. Zhou, Z. Y. Xu and X. Q. Tang,* "New method for determining optimal number of clusters in K-means clustering algorithm", Computer Engineering and Applications, **vol**. 46, no. 16, pp. 27-31, 2010.

[9]. *X. J. Tong, F. R. Meng and Z. X. Wang,* "Optimization to k-means initial cluster centers", Computer Engineering and Design, **vol**. 32, no.8, pp. 2721-2723, 2011.
[10]. *M. Huang, Z. S. He and X. L. Xing and Y. Chen,* "New k-means clustering center select algorithm", Computer Engineering and Applications, **vol**. 47, no. 35, pp. 132-134, 2011.
[11] *L.J. Zhou, H. Wang, W.B. Wang and N. Zhang,* "Parallel K-means algorithm for massive data", Journal of Huazhong University of Science & Technology., **vol**. 40, pp. 150-152, 2012
[12]. *J. H. Song, X. F. Li and Y. J. Liu.* "An Optimized k-means Algorithm for Selecting Initial Clustering Centers", International Journal of Security and Its Applications, **vol**. 9, no. 10, pp. 177-186, 2015.
[13]. *A. K. Jain, R. P. W. Duin and J. C. Mao*, "Statistical pattern recognition: A review", IEEE Transactions on Pattern Analysis and Machine Intelligence, **vol**. 22, no. 1, pp. 4-37, 2000.
[14]. *L. Zhang, Z. W. Shi, J. Wu,* "A Hierarchical Oil Tank Detector with Deep Surrounding Features for High-Resolution Optical Satellite Imagery", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, **vol**. 8, no. 10, 2015.
[15]. *R. Zhang, M. Zhang,* "SAR target recognition based on active contour without edges", Journal of Systems Engineering and Electronics, **vol**. 28, no. 2, pp. 276-281, 2017.
[16]. *A. Eryildirim, I. Onaran,* "Pulse Doppler Radar Target Recognition using a Two-Stage SVM Procedure", IEEE Transactions on Aerospace and Electronic Systems, **vol**. 47, no. 2, 1450-1457, 2011.