Traffic Sign Recognition Technology Based on BOW Model

Hongwei Gao*, Zhe Liu, Yueqiu Jiang
School of Information Science and Engineering,
Shenyang Ligong University.
Shenyang, 110159, P.R. China.
State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences.
Shenyang, 110016, P.R. China.
E-mail: ghw1978@sohu.com

Bin Li
State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences.
Shenyang, 110016, P.R. China.
E-mail: libin@sia.cn

***Abstract*—According to several key technologies in automatic identification technology of traffic signs, this paper makes a detailed study. First of all, from traffic signs segmentation algorithm, a segmentation algorithm based on iterative segmentation and maximum variance between clusters of traffic signs is studied. Secondly, with feature extraction of traffic signs based on SIFT studied, the codebook is generated by these feature clustering and images are described by histograms using Bag of Words (BOW) model. Finally, multi-class classifier based on SVM is designed to classify traffic signs. The experimental results demonstrate the effectiveness and practicality of the BOW model classification algorithm based on the traffic sign images collected in the natural environment.**

***Index Terms*—Image segmentation, SIFT, SVM classifier, BOW model**

I. Introduction

At present, the traffic sign images are identified commonly in these methods: One is the identification to the color and shape features of traffic signs combined with template matching, an image is usually considered as a separate object in these methods which are easily affected by the object angle, light intensity, occlusion of objects and other factors in the process of identifying the target area, and the methods are suitable for a situation of fewer samples. The other is the image feature extraction having been done before the image identified combined with support vector machines and neural network. This method has better robustness for only a small influence of the change of external environment affecting it and is suitable for image classification in the case of more samples. The typical image feature extraction algorithms are Hu invariant moment and SIFT features [1]. Bag of Words algorithm used in the field of image processing is developed in recent ten years and it has been investigated widely[2~7]. Bag of Words model includes image sampling, feature description, building code book and classifier design. Bag of Words model can be imagined as a huge document collection D a total of M documents in D, and each document can be represented as a N vector. Computers are very good at dealing with numerical vector. In recent years, the application of the algorithm in image classification is more and more widely, welcomed by the majority of image processing enthusiasts[8-9].Therefore, this paper carries out research around the identification of several typical traffic signs. Traffic signs segmentation algorithm used to separate sign from the background is studied based on iterative segmentation and maximum variance between clusters. Then with SIFT features of sign extracted, the codebook is generated by these feature clustering and images are described by histograms using BOW model and traffic signs classifier is designed based on SVM at the same time. The experimental results based on the traffic sign images collected in the natural environment demonstrate the effectiveness and practicality of the BOW model classification algorithm.

II. Pretreatment of Traffic Signs

A.Introduction of traffic signs

Traffic signs are specific information transferred to traffic participants with graphic symbols, color and script. They are a kind of auxiliary facilities for traffic command used in the management of traffic and indicating direction in order to ensure the safety and smooth flow of the road. Like other traffic signals, traffic signs play the role of a road language. China's current guidelines for the use of traffic signs are the GB5768-99 standard modified in 1999. Traffic signs divided into two categories include the main signs and auxiliary signs, where the main signs

Manuscript received January 12, 2013; revised April 10, 2013; accepted April 18, 2013.
Corresponding author, Hongwei Gao.

doi:10.4304/jsw.8.10.2652-2659

are divided into 6 subcategories: Warning signs, prohibition signs, indication signs, directional signs, tourist attractions signs and road construction safety signs, a total of 146 kinds of 316 graphic symbols; The auxiliary flag has 5 kinds of 16 graphic symbols. China's traffic signs related directly with road are divided into 3 subcategories, a total of 116 species, 42 prohibition signs, 29 indication signs, 45 warning signs. Figure.1 shows 3 kinds of typical traffic signs.

Figure 1. Some typical traffic signs

B.Pretreatment of Traffic Signs

The maximum between-cluster variance method known as Otsu method [10], the method of threshold selecting was first proposed by Ostu and has been recommended by the majority of scholars because of its simple calculation and good robustness. The Ostu is a method choosing the threshold value automatically instead of settings parameters artificially and it not only applies to the single threshold selection of two areas, but can also be expanded to the multi-threshold selection in multiple regions.

The principle of the Ostu algorithm is derived based on the judgment and analysis on the principle of least squares，and the traffic sign image is divided into target part and the background part by threshold T according to the gray levels. Assume the level of the original grayscale image is L and the number of pixel points with I gray levels is n_I ,I lying in [0，L-1]，so the total image pixels N are：

$$N = n_0 + n_1 + n_2 + \ldots + n_{L-1} \tag{1}$$

The normalized histogram, then

$$p_i = \frac{n_i}{N}, \sum_{i=0}^{L-1} P_i = 1 \tag{2}$$

The area ratio of the target region

$$\theta_1 = \sum_{j=0}^{T} \frac{n_j}{N} \tag{3}$$

The area ratio of the background region

$$\theta_2 = \sum_{j=T+1}^{L-1} \frac{n_j}{N} \tag{4}$$

The average gradation of the target region

$$\mu_1 = \frac{1}{\theta_1} \sum_{j=0}^{T} f_j \times \frac{n_j}{N} \tag{5}$$

The average gradation of the background region

$$\mu_2 = \frac{1}{\theta_2} \sum_{j=T+1}^{L-1} f_j \times \frac{n_j}{N} \tag{6}$$

The average gray of the whole traffic sign image μ is

$$\mu = \mu_1\theta_1 + \mu_2\theta_2 \tag{7}$$

Variance between regions：

$$\sigma_B^2 = \theta_1(t)\theta_2(t)[\mu_1(t) - \mu_2(t)]^2 \tag{8}$$

It is considered to be the best separation state when the variance between two segmented regions is the maximum , and thus the threshold value is T_m determined.

$$T_m = \max[\sigma_B^2(t)] \tag{9}$$

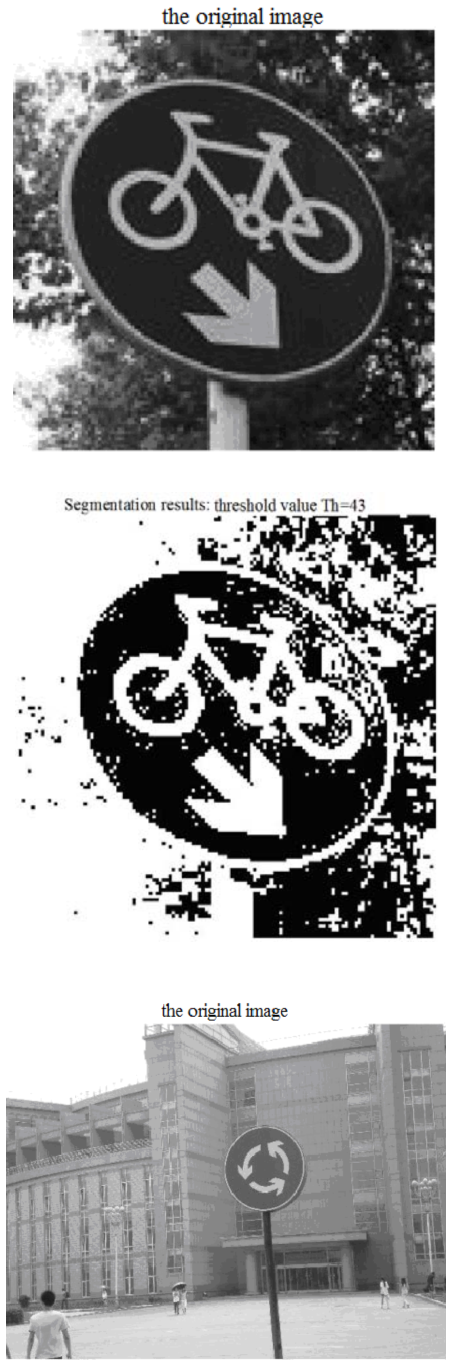
The best discrimination of between-class separability of Ostu equals to exit minimum within-class variance or maximum between-cluster variance in the statistical sense (Equation 9). T_m is the optimal threshold for image segmentation. Variance can reflect the status of the image gray distribution well, the greater the difference of the target area and the background area, the greater the variance. From the traffic signs (Fig. 2) segmented using Ostu, it can be known that the result is bad when the size ratio of the traffic signs and the background is very small.

Figure 2. Effect drawings of traffic signs segmentation based on Ostu algorithm.

Grayscale range is selected as the initial threshold value T_0 , and the original image is divided into the target region and the background region according to the initial threshold value T_0 . Then the integration results on the background region and the target region are averaged to obtain a new threshold value, and the image is divided into a target region and a background region with using the new threshold value. This process mentioned above is repeated until the threshold is no longer changed，and it indicates this iterative threshold has been stabilized as the final threshold used to segment the image renewedly . The iteration threshold algorithm can be defined as:

$$T_{i+1} = \frac{1}{2}\left[\frac{\sum_{K=0}^{T_i} h_k \bullet K}{\sum_{K=0}^{T_i} h_K} + \frac{\sum_{K=T_i+1}^{L-1} h_K \bullet K}{\sum_{K=T_i+1}^{L-1} h_K}\right] \tag{10}$$

Where [0, L-1] is the range of image gray levels, i is the number of iterations, and h_K is the number of points. Iteration will has been finished when $T_{i+1} = T_i$, and then T_{i+1} is the threshold value.



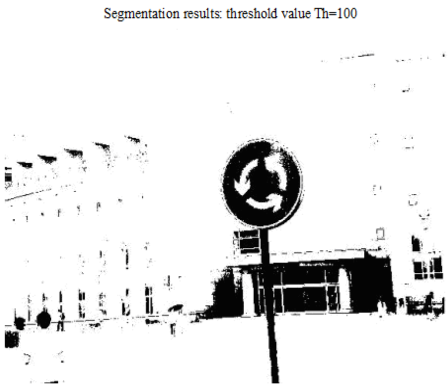


Figure 3. Segmentation effect diagram of the traffic sign image based on iterative threshold

III. BOW MODEL

Bag of Words (BOW) model is firstly applied in text classification and then introduced to the vision field. The main idea of the algorithm is as follows. First, images are regarded as a collection of Independent local features of the image blocks, and then descriptors are generated by image features, finally, the image is decrypted by the descriptors of image blocks and classification is also carried out.) The process of BOW algorithm mainly includes: feature points extraction, feature descriptions, and the generation of the code book and the classifier design. The feature points extraction, characterization and classifier design have been discussed above. The codebook design will be discussed in the next step, the schematic diagram of BOW algorithm is shown in Figure 4.

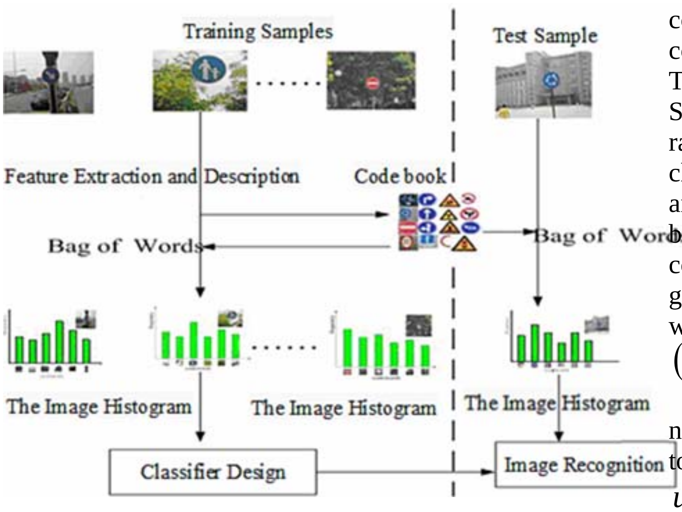


Figure 4. The graph of 0 degree rotation process results

A. SIFT feature extraction

There are some the common feature extraction methods such as intensive sampling, random extraction, and feature point-based block. Feature points can be found in DOG image scale space. First, the scale space is built, difference of scale space is generated by convoluting different scale Gaussian and differential nuclear images[11].

$$D(x,y,\sigma)=(G(x,y,k\sigma)-G(x,y,\sigma))\otimes I(x,y)=L(x,y,k\sigma)-L(x,y,\sigma) \quad (11)$$

Formula (11): $G(x,y,K\sigma)$ and $G(x,y,\sigma)$ are Gaussian functions which scales are variable. $I(x,y)$ is a function which represent image. DOG algorithm is simple which is LOG approximate scale normalized. The extreme points can be seen in DOG after the scale space is built. The information of a location, orientation, scale can be contained in each detected Extreme point. Next, the feature points extracted will be described, and this paper choice SIFT algorithm. SIFT feature is stable to the local features of an image, rotation, scale, zoom, brightness variation, and the viewing angle changed, the affine transformation, the noise is also to maintain a certain stability. The SIFT calculated is usually selected feature point around the 4 × 4 image blocks, and each block is computed on the eight directions on the respective Gaussian accumulated and so each SIFT descriptor is 128 dimensions. Each image block is the 4 × 4 pixel blocks in the extreme point scale. The starting point of the arrow represents the detected feature point. The direction of the arrow represents the direction of feature point. The length of the arrow represents the magnitude of the gradient of the feature point in Figure4 a traffic sign image.

B.The generation of the codebook

The process of building image codebook is as follows. First, image features are extracted with SIFT algorithm, and then the descriptors are generated by the extracted SIFT features .The codebook is generated by SIFT features extracted from the image[12-13]. Codeword is component in the code book. The codeword can be considered a representation of analogous the image block. The method of codebook is generated as follows. First, all SIFT vectors of the image are clustered by K-means, and randomly selected number of feature points are the cluster center in the process of the K-means clustering, and then a number of new cluster centers are generated by a series of iterations. The cluster centers are considered the required code word when the algorithm gradually is Stable. Stable cluster centers make values which are obtained by the formula (11) the minimum. $(x_1,x_2,\cdots,x_n)$ are all the SIFT descriptors. K is the number of cluster centers. $S=\{S_1,S_2,\cdots,S_K\}$ belongs to the subset of a description of each cluster center, and u_1 is each cluster center.

$$\arg\min_S \sum_{i=1}^{K} \sum_{x_j \in S} \left\| x_j - u_i \right\|^2 \quad (12)$$

The number of cluster centers K is the size of the codebook. It is important to choose the size of the codebook. The characteristics of the image can not be contained and can not be represented if codebook is too small, and some noises and the characteristics of the nontarget area can easy to be contained if codebook is too large. Too large or too small will affect the recognition rate of the image. It should be selected according to the specific circumstances of the appropriate size codebook. Image for each image block map to the nearest codeword by the distance of the calculated and each code word in the code book and its. The image is represented with a histogram, and the height of each column is the image the image blocks in the code word on the frequency.

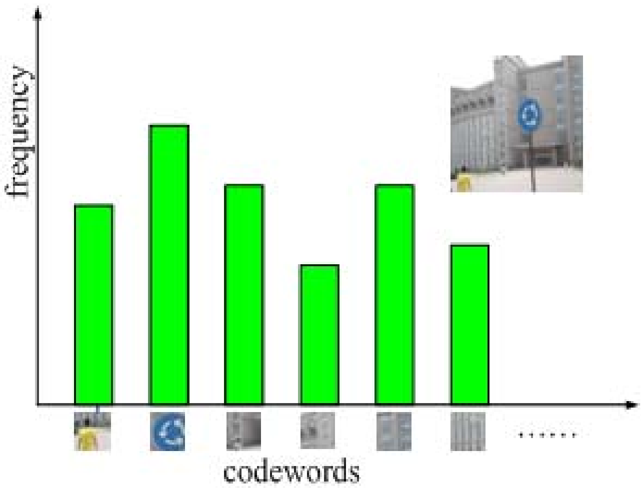


Figure 5. The histogram of traffic signs

Codebook generated is introduced in detail in the Bag of Words model. In the process of codebook model formed in the Bag of Words, codebook is generated by clustering algorithm after SIFT features are extracted in traffic signs. The size of the cluster center represented the size of the codebook. Bag of Words model is originally used in the classification areas. The test sample images generally contain the background and multiple targets. Codebook is more than one target area clustering collection. In this experiment, the test samples are obtained after segmentation traffic signs. Only some noise and destination areas are contained in the image, and the noise region with respect to the target area is negligible. It can be considered that the image contains only a region that the contents of the traffic signs. Codebook formed in the experiment is performed on the image to SIFT features extracted ideal match point. The schematic diagram is shown in Figure 6.

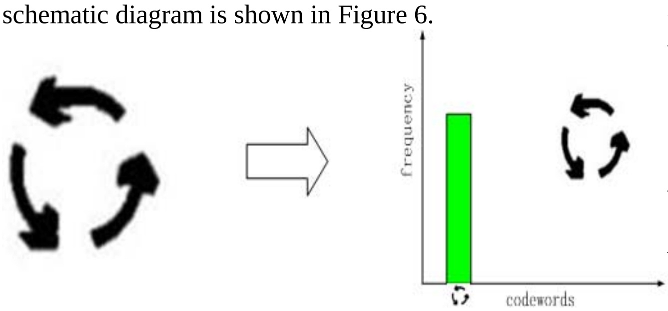


Figure 6. The divided traffic signs and codes

The image of figure 6 contains only traffic sign image that is to say only one codeword in the codebook. There is only one column in the histogram. A traffic sign in the image, the image block and the distance to its own code word map own codeword.

C. Classification Design

In many practical applications, the multi-class classification problems will be faced such as traffic sign recognition in this article. There are 12 classes of the sample of the types of traffic signs. The ordinary support vector machines are put forward for binary classification algorithm. The department is unable to meet the requirements of the classification. How the two types of classification extended to multi-class classification is an important one support vector machine. There are two ways to achieve multi-class classifier design method based on support vector machine. One way is achieved with multiple types of classifiers, and another way is directly designed by using support vector machine itself to transform, it is extended to multi-class classification of multi-class support vector machine. The multi-class classification problems are assumed with K categories $S=\{1,2,\cdots,K\}$, training sample is $\{(x_i,y_i), i=1,2,\cdots,l\}$ $y_i \subset S$,the function multi-class classifier directly designed to be achieved as follows:

$$\min \quad \phi(w,\xi)=\sum_{i=1}^{K}\|w_i\|^2+c\sum_{i=1}^{l}\sum_{j=y_i}^{K}\xi_i^j$$

Constraints:

$$\langle w_{y_i}\bullet x_i\rangle+b_{y_i}\geq\langle w_j\bullet x_j\rangle+b_j+2-\xi_i^j \quad (13)$$

Formula(13)

$\xi_i^j\geq 0, i=1,2,\cdots,l,\ j=1,2,\cdots,K,\ j\neq y_i$

The final decision function is as follows:

$$f(x)=\arg\underset{K}{Max}\left[\sum_{i=1}^{l}\sum_{j=1}^{K}a_i^jK(x_i,x)-\sum_{i=1}^{l}a_i^nK(x_i,x)+b^n\right] \quad (14)$$

The design of multi-class classification Multiple is achieved by types of classifiers. There are mainly three methods which are as follows:

(1) One-against-rest method

One-against-rest method is a way that the K SVM sub-classifiers is constructed. Each sample is a class. The functions were calculated for each sub classifiers on the test data, the output value corresponding to the maximum of the classifier is the category test category.

(2) One-against-one method

One-against-one method is a way that each two types are constructed in the multi-class classifier, and K (K-1) / 2 binary classification device should be constructed if there are K categories. When tested, K (K-1) / 2 sub-classifier be respectively tested, and the highest score classification category is selected to correspond to the test sample category.

(3)M-ary classification method

M-ary classification is classification problem that multiple categories are combined into multiple binary based on the intrinsic relationship between the categories. When tested, the multi-class problem stepwise refinement to a second-class Classification, and all categories eventually are separated, the schematic diagram is as shown in Figure 7.

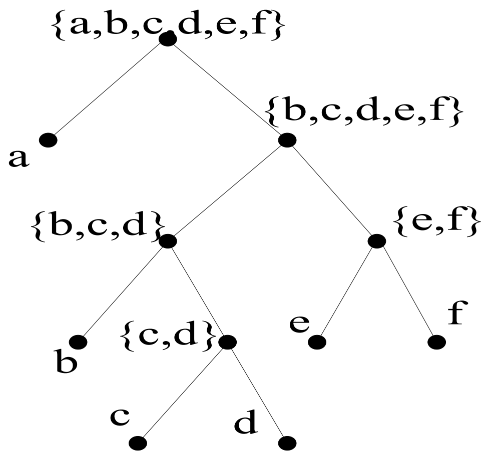
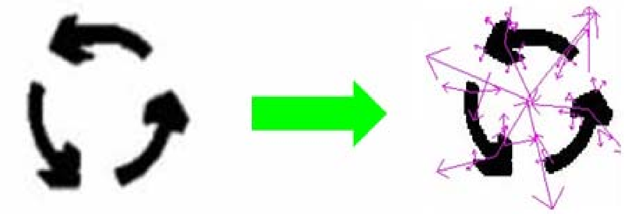


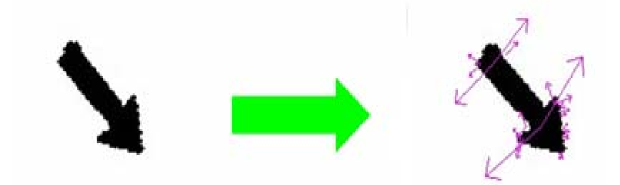
Figure 7. Two binary tree multi-class classification problem is decomposed into two kinds of problems

IV. EXPERIMENTAL RESULTS AND ANALYSIS

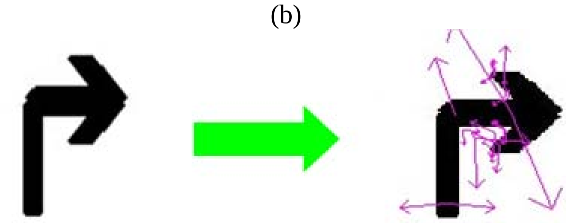
The images used in the experiment are traffic signs that acquired under the natural conditions of the unmanned vehicles and used in traffic signs scene graph segmentation. Platform used in this experiment is to combine OpenCV in VC++6.0 development environment. Experiment the results are shown in Figure 8.



(a)



(b)



(c)

Figure 8. SIFT feature extraction

The 12 types of common traffic signs are selected in this paper as the objects studied. Each traffic sign is rotated with 0 °, 30 °, 60 °, 90 °, 240 ° rotation ,there are a total of 60 traffic signs (12 traffic sign images rotated are not listed in below),as the training samples. In order to facilitate the research, these 12 types of traffic signs are translated into two categories. One is used for the classifier training and learning, another kind is mainly used for testing, training and learning of classifier design.

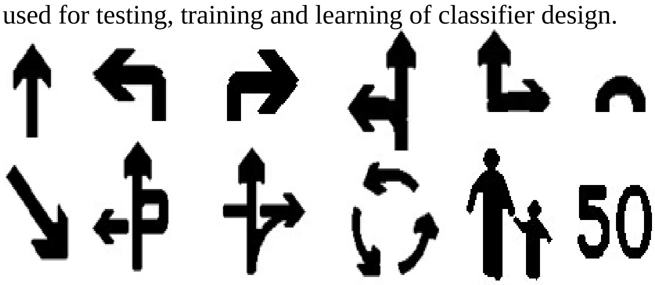


Figure 9. Figure of traffic signs of training samples

In order to test the recognition with support vector machine, the test samples used in this experiment are the traffic signs obtained after image segmentation, so that it can be more close to the actual. A total of 15 pieces of traffic signs are extracted by SIFT feature extraction, and the traffic signs are obtained after the images are the segmented and the 3 types of traffic signs are rotated with 0 °, 30 °, 60 °, 90 °, 240 ° rotation in Figure 9. The 15 pieces of traffic signs are classified and distinguished with BOW model.

(a)、Drive by right

(b)、Pedestrian

(c)、Speed limit

Figure 10. Test samples

In order to test the performance of the algorithm in this paper, Hu invariant moments and the recognition method based on BOW model of this paper are introduced for comparative study.

TABLE I

The average Hu moment invariant features after test samples have been take logarithm

	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6
Drive by right	1.3824	9.4722	14.7238	12.9234	28.5324	19.2503
Pedestrian	1.5603	10.4721	16.0021	14.4538	29.8754	20.2533
Speed limit	1.3213	9.9901	13.2104	10.9703	28.1233	16.1530

Features extracted from the 15 traffic signs are used as a test set with the 60 traffic signs mentioned earlier used as a training set. The recognition effect based on traffic signs has been tested 20 times repeatedly and the recognition rate for each test result is accumulated averaging in order to get the average recognition rate of different algorithms, as shown in Table 2.

TABLE II.

Average recognition rate of traffic signs

	Drive by right	Pedestrian	Speed limit
BOW	96.6%	94.1%	95.8%
Hu invariant moment	93.6%	93.1%	92.8%

There are different degrees of deviation between two recognition algorithms in table 1. It can be understood the extraction process of the traffic signs, traffic signs defaced on the surface, the light changes, signs of deformation and interference factors have had a certain impact on the segmentation of traffic signs, making SIFT features some error, eventually leading to a classification error. The result is satisfactory that traffic signs in using the Bag of Words model combined with support vector machine classification in table 2.

V. CONCLUSION

Automatic recognition algorithms are designed for common traffic signs and the related experiments are made in this paper. First, the traffic signs segmentation algorithm based on iterative segmentation and the maximum between-class variance are studied from traffic signs segmentation algorithm; Second, the characteristics of the traffic signs based on SIFT extraction algorithm is studied. Taking into account the common support vector machine algorithm is proposed for small samples and the Department is unable to meet the requirements of the classification designed for this multi-category classification of traffic signs. The BOW model generated by traffic signs SIFT feature are introduced in detail to prove the effectiveness and practicality of the BOW model classification algorithm based on the experimental results of the traffic sign images collected in the natural environment.

ACKNOWLEDGMENT

This work is supported by China Postdoctoral Science Foundation funded project (Grant No.2013M530954), CALT Innovation Foundation, State Key Laboratory of Robotics Foundation, Shenyang Institute of Automation, Chinese Academy of Sciences（Grant No.2012017）, and also supported by the Liaoning Province Educational Office Foundation of China (Grant No.L2011038).

REFERENCES

[1] P Sompoch, C Xiaoyangv, K Michiro. “Automatic Recognition and Location of Road Signs from terrestrial color Imagery.” Geoinfomatics&DMGIS 2001, Bangkok. Thailand. 2001,pp:1-2.

[2] Man Luo，Miao Qi, “A New Method for Cartridge Case Image Mosaic,” Journal of Software, vol.6, pp.1305-1312, 2011.

[3] Lowe D. “Object recognition from local scale-invariant features,” Proc. the Seventh IEEE International Conference on Computer Vision. IEEE Press, 1999, pp.1150-1157.

[4] Lowe D, David G.”Fitting parameterized three-dimensional models to images,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.13, pp.441-450, 1991.

[5] Zhong W, Qifa K, Michael I, Jian S. “Bundling features for large-scale partial- duplicate web image search,” IEEE Conference on Computer Vision and Pattern Recognition, Miami, 2009, pp.25-32,2009.

[6] Yang Dan, Li Bo, Zhao Hong. “An Adaptive Algorithm for Robust Visual Codebook Generation and Its Natural Scene Categorization Application,” Journal of Electronics & Information Technology. vol.32, pp. 2140-2144, 2010.

[7] Li Fei-fei, Perona PA. “Bayesian hierarchical model for learing natual scene categories,” IEEE Conference on Computer Vision and Pattern Recognition, San Diego, pp. 524-531, 2005.

[8] Jun Zhi, “Design of Simulated Pattern Painting Based on Image Segmentation and Recognition Method,” Journal of Software, vol.6, pp.2232-2238, 2011.

[9] Wei Zhan, Zhiqing Luo, “Research of Vehicle Type Recognition System Based on Audio Video Interleaved Flow for Toll,” Journal of Software, vol.7, pp.741-744, 2012.

[10] Liu Si-ping. “Detection and Recognition Technology for Road Traffic Signs.,”Master Degree Thesis, 2011

[11] He You-song, Wu Wei, Chen Mo, Yang Xiao-ming, Luo Dai-sheng. “Vehicle Image Recognition Study Based on Bag of Features,” Video Engineering,vol.12, pp.106-109, 2002.

[12] Kim S, Kweon IS. “Simultaneous classification and visual word selection using entropy-based minimum description length,” IEEE International Conference of Pattern Recognition, Hong Kong, pp:650-653,2006.

[13] Zhong W, Qifa K, Michael I, Jian S. “Bundling features for large-scale partial- duplicate web image search,” IEEE Conference on Computer Vision and Pattern Recognition,Miami, pp: 25-32,2009.

Hong-Wei Gao, received his Ph.D. degrees for Pattern recognition and intelligent system from Shenyang Institute of Automation (SIA), Chinese Academy of Sciences (CAS) in 2007. Since September 2009, he has being an associate professor of Shenyang Ligong University. Currently, he is the leader of academic direction for optical and electrical detect technology and system. His research interests include digital image processing and analysis, stereo vision and intelligentcontrol.

Zhe Liu, born in 1987, is a postgraduate of University of Shenyang Ligong. His main research interests include image processing.

Yue-Qiu Jiang, received her Ph.D degree in Computer application technology from Northeastern University in 2004. Since 2010, she has been a full professor of Shenyang Ligong University. Currently, she is leader of subject direction for signal and information process. Her research interests include network management and image processing.

Bin Li received the MS degree in Medical Science from China Medical University, Shenyang, China, in 1988. Since 1989, he has been with Shenyang Institute of Technology, Shenyang, China, where he is currently a Professor of the State Key Laboratory on Robotics. His research interest is in rescue robotics and biorobotics.