J. Basic. Appl. Sci. Res., 3(2s)77-80, 2013
© 2013, TextRoad Publication

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

A New Codebook for Object Category Recognition using SIFT

Sonia Gharibzadeh[1], Behzad Mohammadi Alasti[2], Mehdi Abbasgholipour[2]

[1]MSc in Mechatronics Engineering, Islamic Azad University of Ahar, Ahar, Iran,

[2] Assistant Prof. Department of Mechanical Engineering, Islamic Azad University of Bonab, Bonab, Iran

Received: June 10 2013
Accepted: July 10 2013

ABSTRACT

In this paper a new algorithm for codebook construction is proposed. The codebook has an important task in object category recognition when bag of visual word is used for image representation. Traditional codebook construction methods uses only SIFT descriptor. Hence, they considers only visual similarities. The proposed algorithm considers both keypoint location and category information in addition to SIFT descriptors. The new codebook can be implemented efficiently by k-means clustering. Experimental results confirm the performance of the proposed codebook.
KEYWORDS: codebook, SIFT, category recognition,classifier.

1. INTRODUCTION

The Object category recognition is an active field in computer vision society [1]-[4] that deal with training a classifier to recognize the category of object. It categorizes the image based on its semantic content to access visual information on the level of objects (watch, airplain, etc.). There is many application such as worldwide image indexing that deal with thousands of images without any information about the image content. Hence, object category recognition as an automatic tools significantly helps image categorizing and more image indexing [5]. It can be done by bag of words model, parts and structure model, discriminative method and combined recognition and segmentation. Image representation by Bag of word (BOW) model can be applied for classification. BOW is an useful tool for image classification and retrieval. It deals with the image features as words. For classification of a document the BOW is the histogram of word while for classification of images BOW is a bag of visual words. In the other word, one image can be represented by a histogram of the number of visual words occurrences (analogy to words in text documents) [6]. The visual word is a vector of local image descriptor. Scale Invariant Feature Transform (SIFT) is one the best descriptors. SIFT descriptors, represent an image by a set of 128-dementional vectors [7], [8].

To classify the image a codebook (dual of dictionary in a document) is needed. The codebook is a set of codewords. These visual words help to bridge the semantic gap between the low-level image features and the high-level vision. A traditional method to construct the codebook is k-means clustering over all 128-dimentional vectors of training images [5]. The descriptors vector quantize into clusters which are the visual words. In this way, the codewords are cluster centers. All codewords construct the codebook and the number of clusters is the codebook size. Usually, codebook is constructed by only considering the visual similarities of image features. [6] use the category information in addition to the SIFT descriptors. Hence, the codebook is category sensitive by using this algorithm. Traditional codebook construction ignores the spatial relationships among keypoint location(spatial information), which may be helpful for category recognition. Therefore, we proposed to use both keypoint location and category information to construct more discriminative codebook. The rest of this paper is organized as follows: Section2 focuses on the BOW model and traditional codebooks. In section3 category sensitive codebook [6] is discussed . The proposed algorithm is presented in the next Section. Section 5 focuses on experimental results and analyzing of them. Finally, the paper is concluded in section 6.

2. Traditional Codebook

Traditional methods use only SIFT descriptors to construct the codebook [8],[9]. Therefore, only visual information of the image is considered. First of all, the SIFT descriptors are extracted from the training images. Each descriptors is a 128-dementional vector. An image is represented by lots of these vectors. Since, the category of object is the goal, a tool to describe the image category is needed. One can quantize the numerous SIFT vectors in the feature space to some representative vectors [5]. To do this many clustering algorithms can be used. k-means clustering is widely used to vector quantize SIFT descriptor. The centers of the clusters are considered as the codewords and a set of codewords make the codebook. Suppose that $V = (v_1, v_2, ..., v_n)$ are the observation set where v_i is i^{th} 128-dimentional SIFT descriptor. Vector quantization is down by clustering the observations to $Cb = (cw_1, cw_2, ..., cw_k)$ set. Where $k < n$ and cw_i is the i^{th} cluster center or codeword. Each cw_i represents its cluster regio S_i . The objective function to minimize is the sum of squared Euclidian within class distance [5]

$$\underset{S}{argmin} \sum_{i=1}^{k} \sum_{v_j \in S_i} \| v_j - cw_i \|^2 \qquad (1)$$

Corresponding Author: Sonya Gharibzadeh, MSc in Mechatronics Engineering, Islamic Azad University of Ahar, Ahar, Iran, s-gharibzadeh@iau-ahar.ac.ir

Codebook can be viewed as the dictionary. So, the number of occurrence of visual words is calculated for each image. This histogram is sent to classifier to detect the category of the images. The codebook construction has an important step in BOW model and has stimulated some interest [9]-[12]. These works are mainly about how to quantize the feature vector and the selection of more discriminative visual features.

3. Category Sensitive Codebook

As mentioned above , codebook in BOW model is typically constructed by only measuring the visual similarity of image features. Hence, the resulting codebooks may not contain the desired information for object category recognition. Zhang et.al [6] proposed to consider the category information as an additional term into the traditional visual-similarity-only based codebook. Suppose that cat_i is the label of SIFT descriptor v_i. This label indicates from which category of images the feature is extracted. cat_i is a *C*-dimensional vector of binary digits 0 and 1. *C* is the number of image categories. If a SIFT descriptor is extracted from one category of images, the corresponding dimension of cat_i will be assigned the label of 1 and the other dimensions of cat_i will be assigned the label of 0. By defining $vcat_i$ as the extension of 128-dimentional vector v_i by extra dimension αcat_i (α is a weighting vector) the objective function to minimize is as follows:

$$\underset{S}{argmin} \sum_{i=1}^{k} \sum_{vcat_j \in S_i} \| vcat_j - cwcat_i \|^2 \qquad (2)$$

Where $cwcat_i$ the is 128+C dimensional vector. After clustering first 128 dimension of cluster centers can be used as the codewords.

4. Proposed Algorithm

Traditional codebook construction is based on SIFT descriptors which are 128-dimensional vectors. In this method the codebook do not have any information about the vector scale, and orientation. Therefore, the codebook contained only visual similarities information. [6] suggest to construct category sensitive codebook by adding category information to the codebook. Both of mentioned methods do not consider the location of keypoints while this information may be very useful. The location of each descriptor vector and the corresponding scale and orientation for typical image is shown in Fig1. As can be seen in this figure, the location of descriptors are not the same and this differences can be used for discriminate task. Some descriptor vector are longer than others. By considering the scale information of keypoints the shorter vector can be weakened. The direction of the vectors and its distribution can be helpful to recognize the geometric shapes in the image.

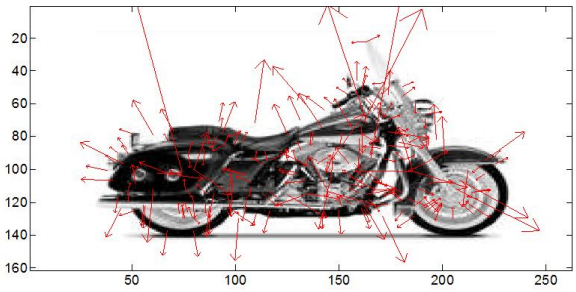

Fig1. Location, scale, and orientation of SIFT descriptor vectors.

The goal of this paper is to propose the algorithm that uses the keypoint location information and simultaneously keep the advantages of traditional and category sensitive codebooks. In this way, the proposed codebook is category and location sensitive and support the visual words. Suppose that $loc_i = (r_i, c_i, s_i, o_i)$ is the keypoint location of 128-dimentional descriptor vector v_i . Where $r_i, c_i, s_i,$ and o_i are row, column, scale and orientation of v_i, respectively. New vector $vloccat_i$ is produced by extending v_i by extra 4-dimenction location vector loc_i and C-dimensional binary vector αcat_i . The 128+4+C dimension vector $vloccat_i$ is sensitive to visual similarity, location and category information. The objective function to minimize is as follows:

$$\underset{S}{argmin} \sum_{i=1}^{k} \sum_{vcatloc_j \in S_i} \| vloccat_j - cwloccat_i \|^2 \qquad (3)$$

Fig.2 shows the proposed procedure to construct the codebook. As can be seen in this figure, location information and binary version of category information is added to 128-dimensional SIFT descriptor vectors. Then, the k-means clustering is applied to the new 128+4+C dimensional vector. First 128 component of cluster centers is considered as the new codebook. The new codebook is sensitive to both category information and keypoint location.

5. Numerical Results

To evaluate the proposed codebook several experiments are made over the Caltech data base. Five images category are adapted: watch, Face, chair, airplane, and Motorbikes. Fig3. shows typical images from these categories. First 50 images of each category is selected to do the experiments. To train the classifier 41 images is used and 9 images is remained for test. To construct the codebook 10 images from each category is used. Then the SIFT is applied to these images. The codebook size for all three method is set to 10. To classify the histogram of codebooks the Bayesian classifier with multivariate normal density and pooled estimation of covariance is used. The k-means algorithm is run 10 times with random initial value and the best result is used. [5] and [6] that introduced above, are implemented for comparison with the proposed codebook.

Table1 indicates to the recognition rate. As can be seen in this table, the performance of three methods for Motorbikes is similar and relatively at high rate . The reason is the simple background of the Motorbikes. With the exception of watch the proposed method is much better than traditional method [5]. That is because of the complicated background of the watch category. Also, the proposed method except for airplane category is better than [6]. In general, the proposed method has the highest average recognition rate compared with [5] and [6], because the proposed codebook is sensitive to both keypoint location and category information as well as visual words. The confusion matrix of the proposed method is shown in Table 2. The diagonal elements of this matrix are equal to the recognition rate. As can be seen in this table, the proposed codebook has the poor performance for the watch category while it has good performance for Face and Motorbikes categories. In general, with simpler background the recognition rate is increased.

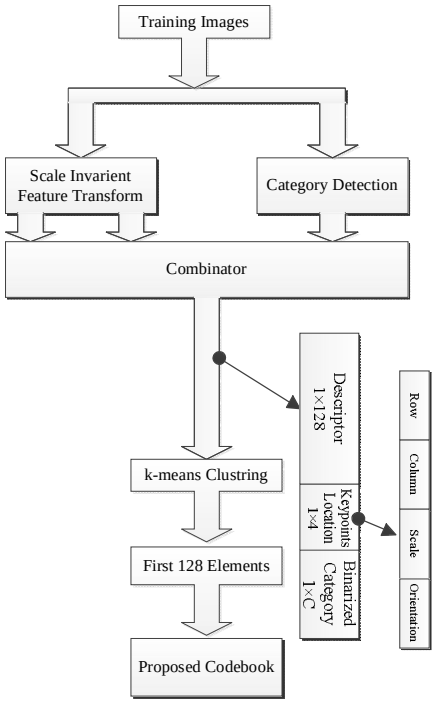

Fig.2 Proposed codebook construction algorithm

Table1 Average recognition rate

Class	[5]	[6]	Proposed
watch	0.6667	0.2222	0.3333
Face	0.5556	0.7778	0.8889
chair	0.3333	0.4444	0.4444
airplanes	0.3333	0.6667	0.5556
Motorbikes	0.8889	0.8889	0.8889
average	0.5556	0.6000	0.62222

Table2 Confusion matrix of the proposed codebook

Predicted / Actual	watch	Face	chair	airplanes	Motorbikes
watch	0.3333	0	0.1111	0	0.5556
Face	0.1111	0.8889	0	0	0
chair	0.1111	0.1111	0.4444	0.2222	0.1111
airplanes	0	0.2222	0.1111	0.5556	0.1111
Motorbikes	0	0	0.1111	0	0.8889

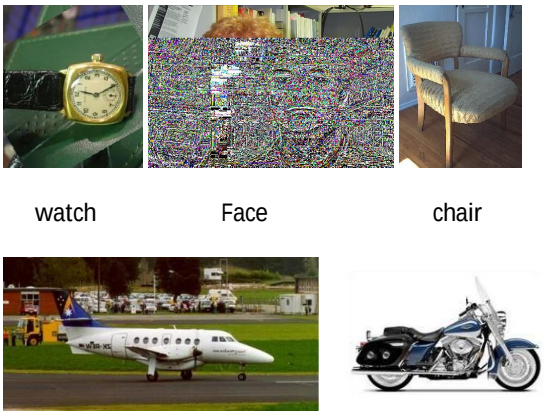

Fig3. Typical images from Caltech database

6. Conclusion

In this paper, new codebook construction algorithms is proposed. The location, scale and orientation of SIFT descriptor vectors is considered to construct the codebook besides the visual words. The new codebook used the category information to make the codebook sensitive to the category. With simple manipulation, the proposed method can be adapted to improve the performance of the traditional codebooks. Experimental results confirm the performance of the proposed method in comparison with both only-visual-similarity codebook and category sensitive codebooks.

REFERENCES

[1] H., Drew S. (2010): Improved machine learning for image category recognition by local color constancy, IEEE international conference on image processing: 3881-3884.

[2] Ommer B., Buhmann J. (2010): Learning the Compositional Nature of Visual Object Categories for Recognition, IEEE transactions on pattern analysis and machine intelligence, 32(3): 501-516.

[3] Sinapov J., Schenck C., Staley K., Sukhoy V., Stoytchev A. (2012): Grounding semantic categories in behavioral interactions: Experiments with 100 objects, robotics and autonomous systems.

[4] Yang L., Jin R., Sukthankar R., Jurie F. (2008): Unifying discriminative visual codebook generation with classifier training for object category, IEEE conference on computer vision and pattern recognition: 1-8.

[5] Sivic J. S., Zisserman A. (2003): Video google: A text retrieval approach to object matching in videos, In Proc. of ICCV, 2: 1470-1477.

[6] Zhang C., Liu J., Quyang Y., Tian Q., Lu H., Ma S. (2009): Category sensitive codebook construction for object category recognition, IEEE international conference on image processing: 329-332

[7] Moreno P., Jimenes M., Bernardino A., Santos J., Blanca N. (2007):A comparative study of local descriptors for object category recognition: SIFT vs HMAX, Pattern recognition and image analysis, 4477:.515-522.

[8] Sande K., Gevers T., Snoek C. (2008):Color descriptors for object category recognition, 4th european conference in colour in graphics, imaging and vision: 378-381.

[9] Hsu Y. W. H., Chang S.-F. (2005):Visual cue cluster construction via information bottleneck principle and kernel density estimation, In Proc. of CIVR.

[10] Lu T. (2010): A survey of VQ codebook generation, journal of information hiding and multimedia signal processing, 1(3):190-203.

[11] Winn K., Criminisi A., Minka T.(2005):Object categorization by learned universal visual dictionary, In Proc. of ICCV:1800-1807.

[12] Perronnin F., Dance C., Csurka G., Bressan M. (2006): Adapted vocabularies for generic visual categorization, In Proc. Of ECCV:464-475.