Statistical Approach for Supervised Codeword Selection

Kihong Park, Seungchul Ryu, Seungryong Kim, and Kwanghoon Sohn
School of Electrical and Electronic Engineering, Yonsei University,
50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea

ABSTRACT

Bag-of-words (BoW) is one of the most successful methods for object categorization. This paper proposes a statistical codeword selection algorithm where the best subset is selected from the initial codewords based on the statistical characteristics of codewords. For this purpose, we defined two types of codeword-confidences: cross- and within-category confidences. The cross- and within-category confidences eliminate indistinctive codewords across categories and inconsistent codewords within each category, respectively. An informative subset of codewords is then selected based on these two codeword-confidences. The experimental evaluation for a scene categorization dataset and a Caltech-101 dataset shows that the proposed method improves the categorization performance up to 10% in terms of error rate reduction when cooperated with BoW, sparse coding (SC), and locality-constrained liner coding (LLC). Furthermore, the codeword size is reduced by 50% leading a low computational complexity.

Keywords: Distinctive codeword selection, object categorization, bag-of-word

1. INTRODUCTION

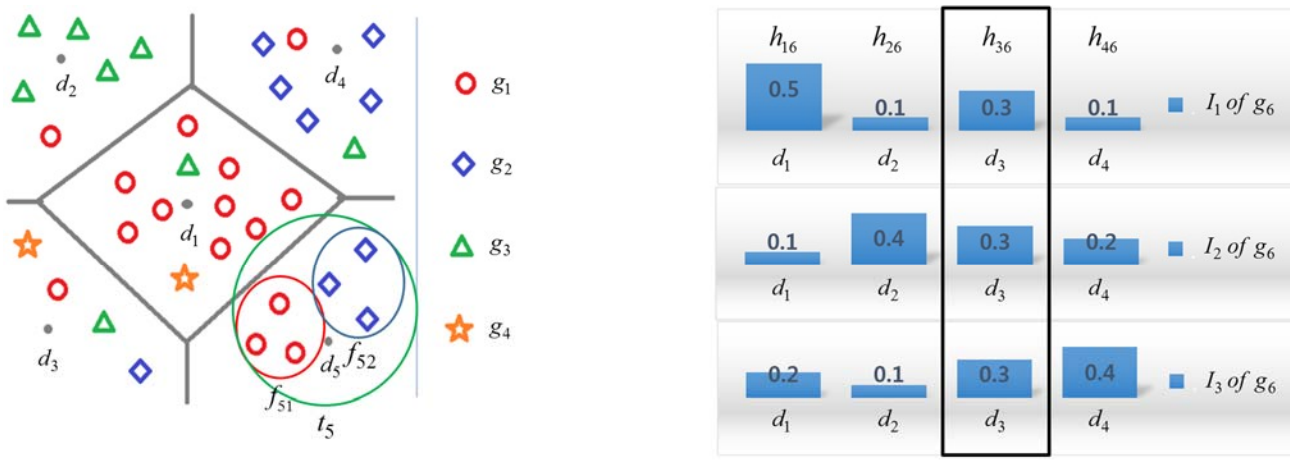
Over the last decades, the bag-of-words (BoW) [1] has received a lot of attentions as one of the most successful object categorization methods. In the BoW, an image is represented by an orderless collection of quantized local features so-called codewords. In the conventional BoW-based algorithms, the codeword size is manually selected by a user and is commonly defined up to tens of thousands for ensuring enough information encoding. However, such a huge size of codewords causes an enormous computational cost [2].

In order to address such a high computational cost problem, codeword selection algorithms have been proposed focusing on eliminating non-discriminative codewords [3]. The typical approaches are based on principal component analysis (PCA) [4-6]. The PCA is one of the most widely used data dimension reduction methods, which determines axes maximizing variance of codewords. However, the PCA-based methods do not consider category information, thus, it cannot select the most informative subset of codewords.

In this paper, we propose a statistical approach for codeword selection. By analyzing feature distributions in the features domain and codeword variances of the BoW histogram domain, we construct the statistical criteria, cross- and within-category confidences. First, the cross-category confidence sorts out codewords by eliminating indistinctive codewords across categories with a probabilistic scheme. Second, the within-category confidence selects codewords by eliminating inconsistent codewords within each category based on the statistics of codewords. Our codeword selection algorithm selects a highly discriminative subset of codewords by combining those two confidences.

The remainder of this paper is organized as follows. In Section 2, we review the related algorithms for the codeword selection. In Section 3, a statistical codeword selection algorithm is described with the cross- and within-category confidences. In Section 4, the performance of the proposed method is evaluated for two public datasets: the scene categorization dataset [7] and the Caltech-101 dataset [8]. Lastly, we conclude the paper with a discussion for future works in Section 5.

Intelligent Robots and Computer Vision XXXII: Algorithms and Techniques, edited by Juha Röning, David Casasent
Proc. of SPIE-IS&T Electronic Imaging, Vol. 9406, 940609 · © 2015 SPIE-IS&T
CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2078771



(a) (b)

Figure 1. Descriptions of (a) cross-category confidence and (b) within-category confidence.

2. RELATED WORKS

2.1 Variants of Bag of Word

Over the past decades, a lot of efforts have been done to improve the original BoW algorithm in object recognition and scene categorization. Lazebnik *et al.* [7] proposed spatial pyramid matching (SPM) strategy to employ geometrical information, and Nowak *et al.* [9] proposed to use a dense sampling method for feature extraction based on the performance comparisons of several sampling methods. Bosch *et al.* [10] introduced a concept as an intermediate features using relations between codewords. Some researches attempted to improve an encoding procedure. Perronnin *et al.* [11] and Yang *et al.* [12] employed a fisher kernel (FK) and a sparse coding (SC), respectively. Wang *et al.* [13] improved SC by using nearest codewords strategy: locality-constrained linear coding (LLC). Such coding methods considerably improve the performance due to the prevention of information loss commonly occurred during the hard voting histogram construction. Unlike these methods, our approach is focused on the codeword selection scheme, which will improve the performance of above methods.

2.2 Codeword Selection Algorithms

A feature subset selection is one of the most important tasks for an efficient object categorization system. By eliminating indistinctive features, it can reduce the computational complexity and increase the categorization precision. According to the adoption of category information, feature subset selection methods are classified into two types: supervised and unsupervised methods.

First of all, unsupervised methods try to reduce dimensions using the distribution of features itself. The PCA is one of the most widely used unsupervised methods taking an advantage of well-defined singular value decomposition (SVD). Jegou *et al.* [4] proposed a compact descriptor based on eigenvalues. Fergus *et al.* [5] drove robust shapes for the part based recognition by PCA. LeCun *et al.* [6] proposed PCA-derived feature to extract pose and lightening invariant object recognition. However, they did not consider for feature noise since they are based on the assumption that all features are noise-free and have consistency across different images.

Secondly, supervised method selects an informative codeword subset guided with label information. Wang *et al.* [3] adopted a boosting algorithm to select discriminative codewords, and Ji *et al.* [2] considered the task-dependent selection and the sparse coding technique for compressing the codebook. Rakotomamonjy *et al.* [14] proposed a support vector machine (SVM) based discriminative codewords selection method.

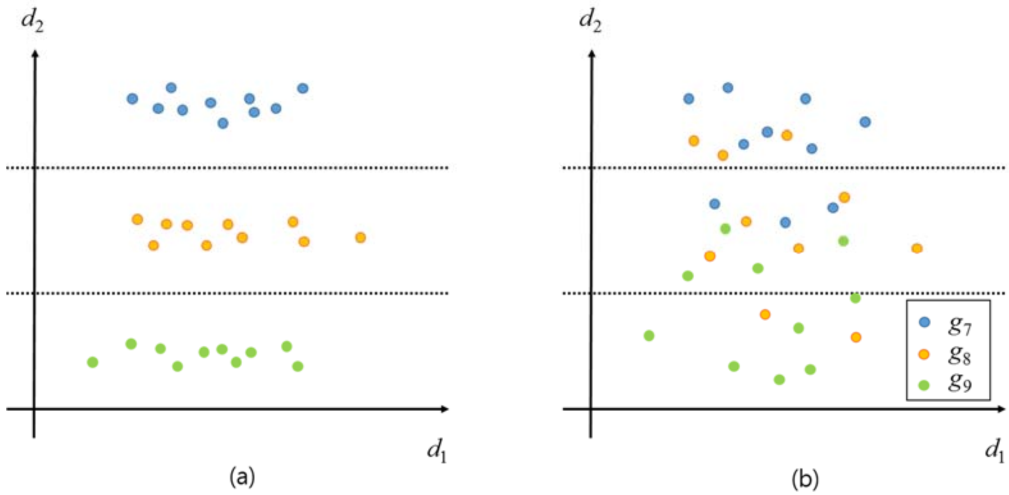


Figure 2. BoW feature distributions followed by variances of a d_2 codeword. (a) low variance of a d_2 codeword. (b) high variance of a d_2 codeword.

3. STATISTICAL CODEWORD SELECTION ALGORITHM

3.1 Problem Statement

The result of the categorization would be successful when each category has a completely distinctive distributions from other categories. In other words, categories having similar histogram distributions may increase ambiguity of the categorization system. Thus, the distribution difference among the categories should be enlarged for the higher categorization performance. To alleviate these problems, we defined a novel codeword selection criteria to eliminate homogeneous codewords and select discriminative codewords. This selected the codeword subset enlarges the distribution difference.

The criteria are based on the statistical information such as variances and distributions of codewords. Since the first category confidence is calculated by analyzing category distributions of the i^{th} codeword, we decide to call the criterion ′cross-category confidence′ and denote by $C_{cross,i}$. In response to the cross-category confidence, the second category confidence analyzing an intra-category variance of the i^{th} codeword is called ′within-category confidence′ and denoted by $C_{within,i}$. Our approach is focused on how to compute these category confidences and eliminate non-informative codewords based on these confidences.

3.2 Cross-Category Confidence

The first category confidence has the highest value when all the features of a codeword show a single category in the feature domain. Since the codeword only exist in histograms of the category images, the histogram distribution differs from other categories, thus the codeword enhances the categorization result. In general, we add up some highest ratios and set the confidence of the codeword, since many codewords consist of more than two categories. It can be shown in Figure 1-(a). In the clustering process, a d_1 codeword consists of features being extracted from g_1 category images more than other category images. In other word, there are many g_1 category images and a few other category images among the training images having the i^{th} codewords. Thus, a query image having d_1 codewords would have a high possibility to be the g_1 category image. It means that the d_1 codeword could be the index of the g_1 category images. On the other hand, a d_3 codeword consists of features being extracting from various categories with the same values, and it may not be the index of a specific category because each category has same probability. Thus, we picked some highest category ratios which are higher than the average in each codeword and add up the highest ratios.

Based on this strategy, the cross-category confidence of the i^{th} codeword $C_{cross,i}$ is represented as follows:

$$C_{cross,i} = \sum_{j=1}^{N_{category}} \max\left(\frac{f_{ij}}{t_i} - a_i \ , \ 0 \right), \tag{1}$$

$$a_i = \frac{1}{e_i} \sum_{j=1}^{N_{category}} \frac{f_{ij}}{t_i} \tag{2}$$

where f_{ij} means the number of training image features which belong to the i^{th} codeword and has the j^{th} category label. $N_{category}$ is the number of categories and t_i is the total number of features constructing the i^{th} codeword. a_i is the average of all category feature ratios which form the i^{th} codeword. We add up the feature ratios of all categories and divide them into the number of the categories which exists in the i^{th} codeword e_i .

3.3 Within-Category Confidence

The second category confidence is inversely proportional to a histogram value variance of a codeword among the same category images. As we pointed out before, each codeword consists of features coming from several categories. It means that different category images can have similar kinds of codewords and the categorization should be based on histogram values of codewords. In this situation, a high variance histogram value of a codeword interrupts categorization process as shown in Figure 2. Figure 2-(a) and (b) have the same histogram values in the BoW histogram domain except a d_2 codeword. Because the histogram value of the d_2 in Figure 2-(b) has the high variance, histograms of different category images are mixed in boundaries. It makes the support vector machine (SVM) difficult to classify categories. Thus, we conclude that low variance codewords at the BoW histogram domain are stable to be classified. It can be observed in Figure 1-(b). d_3 histogram values of g_6 category images show a consistent value while other codewords are not. If the d_3 histogram value of a query image is '0.3,' the image could be the g_6 category because many g_6 category images have the d_3 histogram value of '0.3'. Based on this concept, within-category confidence $C_{within,i}$ is obtained by the following formula :

$$C_{within,i} = 1 / \sum_{j=1}^{N_{category}} \mathrm{var}(h_{ij}) \ , \tag{3}$$

where h_{ij} is the i^{th} codeword value of each image belonging to the j^{th} category in the BoW histogram domain. $\mathrm{var}(h_{ij})$ is a variance of the h_{ij} .

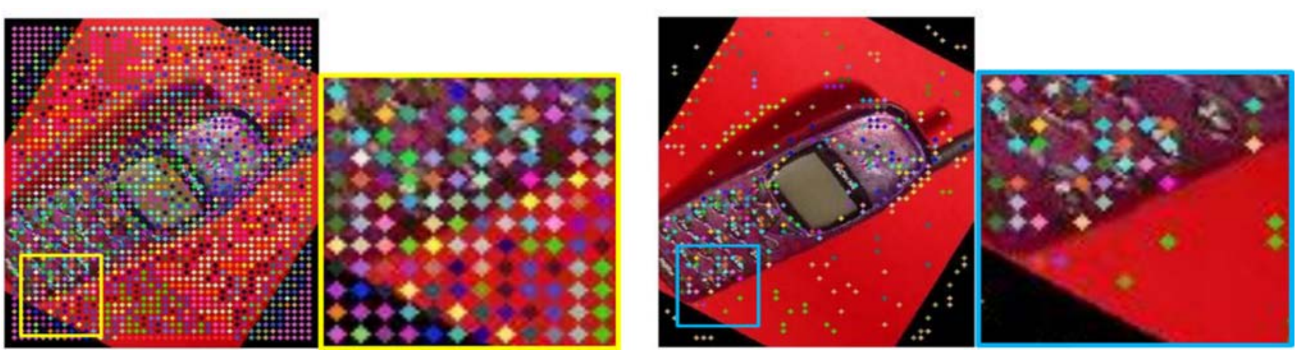


Figure 3. Elimination of error codewords in a real image.

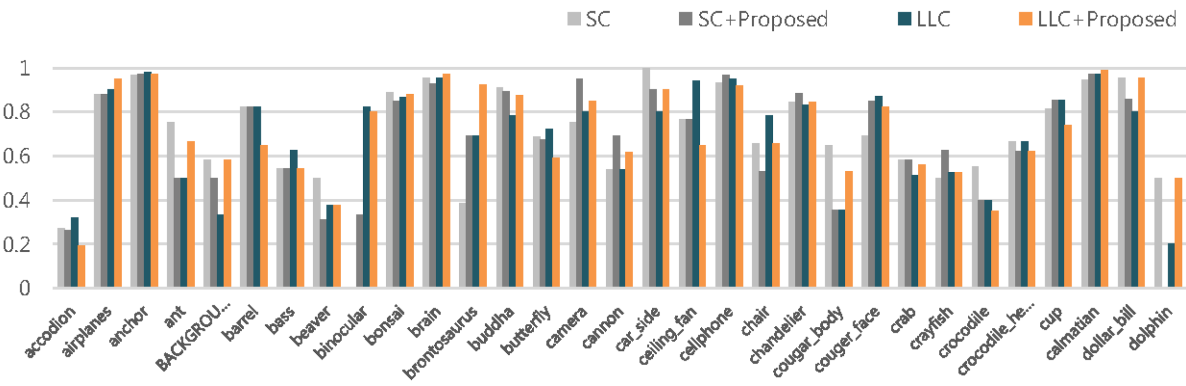


Figure 4. Individual effects of categories on the state of the art algorithms.

3.4 Overall Confidence

Even if both confidences enhance the distribution differences between the categories, they try to solve different aspects. The cross-category confidence is concentrated on the kinds of codewords while the within-category confidence is focused on the differences between the numbers of the same codewords. Therefore, both confidences enhance the categorization performance individually, and complement each other at the same time. We synthesize these two confidences and select reliable codewords. As cross- and within-category confidences are higher, the reliability is also higher. However, since the ranges of these two confidence values differ from each other, weighting parameter is needed. Total confidence of the i^{th} codeword is shown as follows:

$$C_{total,i} = \alpha \cdot C_{cross,i} + \beta \cdot C_{within,i} \tag{4}$$

where α and β are constant values.

Figure 3 shows the effect of our codeword selection algorithm for a real image. Each color of a quadrangle means a codeword type. In codeword selection process, many codewords coming from homogeneous regions disappeared. Because these homogeneous codewords are distributed in various categories by various values, their confidences are very low and eliminated in the codeword selection process.

4. EXPERIMENTAL RESULTS

We used the scene categorization dataset [4] and the Caltech 101 dataset [8] to evaluate the categorization performance of our algorithm. The scene categorization dataset consists of 15 categories. In experiments, half of the randomly selected images in each category were used for training classifier and the other images are used to evaluate the performance. Caltech 101 dataset consists of 101 categories where each category contains more than 80 images. We used 30 images of each category for training, and the rest of the images were used for testing.

Table 1. The precision comparisons of each step of the proposed method for scene categorization dataset [4].

	BoW	BoW + PCA [1]	BoW + Proposed-Cross	BoW + Proposed-Within	BoW + Proposed-Both
Precision (%)	50.2	59.1	**56.5**	**55.7**	**61.1**

In order to validate the performance of our codeword selection algorithm, BoW [1], SC [8], and LLC [9] are employed as base algorithms. We also compare the proposed method with the PCA-based feature selection which is the most widely used codeword selection algorithm. Table 1 shows the categorization precision of the proposed method compared to the original BoW and BoW + PCA [1]. The results show that the proposed method outperforms both BoW and BoW + PCA methods. The contributions of each step are also analyzed in Table 1. Cross-category confidence and within-category confidence improved the performance by 6.3% and 5.5%, respectively. Combined together, 10.9% of precision improvement was achieved.

Table 2. The evaluation results of the proposed method in terms of cooperation with the SC [8] and LLC [9] for Caltech 101 dataset [8].

	SC [8]	SC [8] + Proposed	LLC [9]	LLC [9] + Proposd
Precision (%)	70.7	**73.1**	71.3	**73.6**

To evaluate the proposed method, we also applied our codeword selection algorithm to two state-of-the-arts algorithms: SC [8] and LLC [9] for Caltech 101 datset. Table 2 summarizes the precision results. Note that ′+ Proposed′ indicates that the proposed codeword selection method was applied. The categorization precisions were improved by 2.1 % and 2.3% for SC and LLC, respectively. Figure 4 shows more detailed results for each category. As shown in Figure 4, for most categories, the performance was enhanced. The experimental results proved that our algorithms may cooperate with other base algorithms well improving the categorization performance.

5. CONCLUSION

In this paper we have proposed the statistical approach for the codeword selection. We defined two codeword confidences such as cross- and within-category confidence for selecting the codeword subset with a low complexity and a high precision. The cross-category confidence sorts out indistinctive codewords across categories. Furthermore, the within-category confidence selects stable codewords from image to image. Experimental evaluations show that the proposed method outperforms the conventional methods and cooperates with other base algorithms well. Nevertheless, our approach may destroy the valuable information during a codeword elimination process. Thus, our future work will focus on eliminating only noise factors from the low confidence codewords to avoid the information loss with a supervised clustering algorithm.

6. REFERENCES

[1] G. Csurka, C. R. Dance, L. F. Jutta, J. Willamowski, and C. Bray., "Visual Categorization with Bags of Keypoints," *International Workshop on Statistical Learning in Computer Vision*, (2004).

[2] R. Ji, H. Yao, W. Liu, X. Sun, and Q Tian., "Task-Dependent Visual Codebook Compression," *IEEE Transactions on Image Processing*, Papers 21(4), 2282-2293 (2012).

[3] L. Wang., "Toward a discriminative codebook: codeword selection across multi-resolution," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2007).

[4] H. Jegou, M. Douze, C. Schmid, and P. Perez., "Aggregating local descriptors into a compact image representation," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2006).

[5] R. Fergus, P. Perona, and A. Zisserman., "Object Class Recognition by Unsupervised Scale-Invariant Learning," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2003).

[6] Y. LeCun, F. J. Huang, and L. Bottou., "Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2004).

[7] S. Lazebnik, C. Schmid, and J. Ponce., "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2006).

[8] L. Fei-Fei, R. Fergus, and P. Perona., "One-Shot learning of object categories," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, Papers 28(4), 594-611 (2004).

[9] E. Nowak, F. Jurie, and B. Triggs., "Sampling Strategies for Bag-of-Features Image Classification," *Proceeding of Ninth European Conference of Computer Vision*, (2006).

[10] A. Bosch, A. Zisserman, and X. Munoz., "Scene Classification Via pLSA," *Proceeding of Ninth European Conference of Computer Vision*, (2006).

[11] F. Perronnin, and C. dance., "Fisher Kernels on Visual Vocabularies for Image Categorization," *IEEE Computer*

Society Conference on Computer Vision and Pattern Recognition, (2007).
[12] J. Yang, K. Yu, Y. Gong, and T. Huang., "Linear spatial pyramid matching using sparse coding for image classification," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2009).
[13] J. Wang, J. Yang, K. Yu, F. Lv, and T. Huang., "Locality-constrained Linear Coding for Image Classification," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2010).
[14] A. Rakotomamonjy., "Variable Selection Using SVM-based Criteria," *The Journal of Machine Learning Research*, Papers 3(1), 1357-1370 (2003).