International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.6 (2015), pp.265-276
http://dx.doi.org/10.14257/ijmue.2015.10.6.26

Large-Scale Image Retrieval with Bag-of-Words and *k*-NN Re-Ranking

Pang Haibo, Liu Chengming, Zhao Zhe and Li Zhanbo

School of Software Technology, Zhengzhou University, Zhengzhou, China, 450002
phb@zzu.edu.cn

Abstract

Image retrieval methods have been significantly developed in the last decade. The BOW (Bag-of-words) model lacks spatial information. Some methods stem from BOW approach which is recently extended to a vector aggregation model. Most of them are either too strict or too loose so that they are only effective in limited cases. In this study, we present a novel feature extraction method for image retrieval. We acquire the gradients features from the p.d.f (Probability density function) because of essentially representing the image. We construct the features by the histogram of the oriented p.d.f gradients via aggregation of the orientation codes. Then, we adopt the PCA (Principal component analysis) method to reduce the dimensionality of BOW. Furthermore, we introduce a novel and robust re-ranking method with the k-nearest neighbors. We estimate our method using various datasets. In the experiments on scene retrieval, the proposed method is efficient, and exhibits superior performances compared to the other existing methods.

Keywords: *Image retrieval; bag-of-words; principal component analysis; k-nearest neighbors*

1. Introduction

Image retrieval has attracted keen attentions in the computer vision domain in the last decade. Most state-of-the-art image retrieval approaches adopt the standard *BOW* [1] model on account of the advances of the local descriptors such as *HOG* (Histogram of oriented gradients) [2].

Image retrieval includes such as object recognition [3-4], and scene retrieval, posing a challenge to cope with significant variations of the objects as well as the change of scene in the image.

BOW is based on the local descriptors densely extracted in an image which are coded into features and come into being as the image feature the histogram of the features. The *BOW* has been recently extended to the methods aggregating vectors [5], *etc.* Those orientation codes are aggregated around respective features into the histograms at last.

Although *BOW* model works generally well, it has two problems: 1) the loss of spatial information when representing the images as histograms of quantized features; 2) the deficiency of feature's discriminative power, either because of feature's intrinsic limitation to tolerate large variation of object appearance, or due to the degradation caused by feature quantization.

In this study, we put forward a novel approach to extract effective features for image retrieval. The similarity measure is aimed to handle object translation, scaling and rotation, and performs well with moderate object deformation.

For describing the images, our proposed approach extends the discrete representation in *BOW* to the continuous *p.d.f*. Since the *p.d.f* essentially represents the image, we extract features from the *p.d.f* adopt the gradients on the *p.d.f* in a manner similar to *HOG* applied to image pixel function. By means of computing the gradients, the mean shift vectors [6]

ISSN: 1975-0080 IJMUE

are naturally induced and those vectors are coded in the light of their orientations.

While the retrieval is performed such as by linear *SVM* (Support vector machine), much research effort has been made to resolve effective feature representation [7].

We aim at raising the retrieval accuracy while maintaining affordable memory and time cost. An image that contains the query target may not be visually close to the query due to feature variations caused by the changes of view point, deformation or occlusion. However, some of query's neighbors, which can be considered as variations of the query object, may obtain the same or similarity features with that image.

Therefore, we put forward a re-ranking method with the *k-NN* (*K*-nearest neighbors) of the query. Localized objects in the top-*k* retrieval images are also used as queries to perform retrieve. A new score of each database image is collaboratively determined by those ranks, and re-ranking is performed using the new scores.

Accordingly, our method can successfully retrieve the objects with large variations, owing to it is rank-order based, which discards their distances and the features when calculating the score.

The remaining part of this paper is organized as follows: Section 2 introduces the related works. Section 3 describes proposed methods, including the details of bag-of-words, PCA and *k-NN* re-ranking method. In Section 4, we introduce datasets and features which our experiments used images, extract features, the experimental results and the performance analysis. Conclusions and suggestions for our future works are given in Section 5.

2. Summary of the Related Works

In this Section, we now review related work in the fields which are closest to our large-scale image retrieval problem. We briefly introduce the methods designed to handle the above mentioned problems of the bag-of-words model.

To alleviate the information loss in feature quantization, soft assignment on features is adopted in [8]. The probabilistic relationships between the features are learned in [9]. Feature metrics are also learned either to reduce the descriptor dimensionality or to increase the feature discriminative power.

In the *Bow* (bag-of-words) framework [10], an image is represented by means of a bag of local descriptors densely extracted in the image, and then is finally characterized by a histogram of features quantizing the underlying probability distribution of the local descriptors. The vocabulary provides a discrete partitioning of the feature space by features.

Improved methods include incorporating contextual information into the vocabulary, building super-sized vocabulary [11], *etc.* Typically, multi-vocabulary merging can be performed either at rank level, or at score level.

Matching refinement feature-to-feature matching is a key issue in the Bow model. To improve precision, some works analyze the spatial contexts [12] of SIFT features, and use the spatial constraints as solution to refining matching.

On the other hand, to address the problem of vocabulary correlation, the literature [13] present to create the vocabularies jointly and decrease correlation from the view of vocabulary generation.

In another way to compensate the deficiency in feature matching is to automatically expand the query [14]. It tends to improve the retrieval performance especially when the appearance of the object has large variation. Though a faster method is proposed recently [15], the re-ranking is still performed only on the top-ranked images.

In [16], pair wise feature distances between images are updated using *k-NN*. However, constructing such pair-wise data structure is computationally with large dataset. We propose a *k-NN* re-ranking method without sacrificing much efficiency.

Many works have proposed to transform high-dimensional vectorial representations into compact codes. This includes *LSH* (Locality sensitive hashing) [17], *SSH* (Semi-supervised hashing) [18], *ITQ* (Iterative quantization) [19].

Although the significant differences between above-mentioned algorithms, all of them include a projection of the original image characteristics into an intermediate real-valued space. The projections are either learned in an unsupervised manner or random (as in *LSH*), for instance with *PCA* (as in *SSH*) or with an algorithm which reduces the quantization error (as in *ITQ*).

3. Proposed Methods

3.1 Oriented Probability Density Function Gradients

Considering an input image, *N* local descriptors, are extracted at dense spatial positions with various scales; it can be denoted by

$$x_i \in \mathbb{R}^d,\ \ i = 1, \cdots, N \tag{1}$$

While the bag of those descriptors has been used to discretely represent the image, we apply kernel density estimator to obtain the p.d.f by formula (2),

$$p_f(x) = \frac{1}{N}\sum_{i=1}^{N} f(\|x - x_i\|_2^2) \tag{2}$$

Where $f(z): \mathbb{R} \to \mathbb{R}$ indicates the differentiable function for kernel; e.g., $f(z) = C_{d,h}e^{-\frac{z}{2h}}$ with the parameter *h*, say *h* = 0.1 in our experiments, and the normalization constant $C_{d,h}$. We use this *p.d.f* for constructing an effective image feature.

The gradients effectively characterize the "shape" of the *p.d.f* from the geometrical viewpoint, as is the case with HOG applied to extract geometrical feature of an image pixel function.

The gradient vector of the *p.d.f* is given by formula (3)

$$\nabla p_f(x) = \frac{2}{N}\sum_{i=1}^{N}(x - x_i)\ f^{'}(\|x - x_i\|_2^2) = p_f(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - x)g(\|x - x_i\|_2^2) \tag{3}$$

Where $g(z) = -2f'(z)$ is improper to straight forwardly aggregate the *p.d.f* gradient vectors themselves. Thus, we consider the orientation coding of the *p.d.f* gradients (3), followed by aggregation into histograms.

The orientation coding is usually applied to image gradients such as in *HOG*. The orientation of the image gradients is coded based on a lot of the bins, forming over complete set to describe any oriented gradients. Then, we employ the complete set of bases given by *PCA*.

3.2 Principal Component Analysis

PCA is applied to the *p.d.f* gradient vectors normalized in unit L_2 norm $\nabla p_f / \|\nabla p_f\|_2$ which indicate only the orientations on a unit hyper sphere. Thereby, we acquire the d orthonormal eigen vectors, $u_j, j = 1, ..., d, u_j u_k = \delta_{jk}$. Along each basis vector, we can take into account positive and negative ones, which totally provides *2d* orientation bins by

$$\begin{aligned} C(v;\{u_j\}_{j=1}^{d}) = [&\max(u_1^T v, 0)^2, \max(-u_1^T v, 0)^2, ..., \\ &\max(u_d^T v, 0)^2, \max(-u_d^T v, 0)^2]^T \in \mathbb{R}_+^{2d} \end{aligned} \tag{4}$$

Where v shows the *d*-dimensional vector to be coded. This coding produces rather sparse orientation codes in which at most *d* components are nonzero, and the code has a unit sum for $\|v\|_2^2 = 1$.

The orientation of the *p.d.f* gradient vector is coded by $C(\frac{\nabla p_f(x)}{\| \nabla p_f(x) \|_2}; \{u_{j=1}^d\})$. For convenience, we can leave out $\{u_{j=1}^d\}$ in the followings.

PCA produces the eigenvalues e_j, and the eigenvalue stands for the power of the code on the corresponding basis $e_j = E_x[(u_j^T \frac{\nabla p_f(x)}{\| \nabla p_f(x) \|_2})^2]$, and thus it is utilized to normalize the orientation codes as in *PCA* whitening:

$$\hat{C}(\frac{\nabla p_f(x)}{\| \nabla p_f(x) \|_2}) = E^{-1} \hat{C}(\frac{\nabla p_f(x)}{\| \nabla p_f(x) \|_2}) \qquad (5)$$

Where $E = diag(e_1, e_2, ..., e_d, e_d) \in \square^{2d*2d}$.

In spite of this weighting, the orientation codes are equally dealt with by enhancing the orientations, but rarely occur while suppressing the common ones that are frequently found on the whole. The rare orientations would be improves the discriminative power.

3.3 Aggregation of *p.d.f* Gradient Orientation Codes

The orientation codes (5) are aggregated around the words which are cluster centers in the local descriptor space $\square^d$. We define the aggregation in the following continuous form with the *p.d.f*:

$$\int W(x,\mu) \| \nabla p_f(x) \|_2 \hat{C}(\frac{\nabla p_f(x)}{\| \nabla p_f(x) \|_2}) dx \qquad (6)$$

$W(x,\mu)$ is the weighting function. To reduce the continuous form into a tractable discrete one, it should be noted that the local descriptors $x_i, i = 1,...,N$ are assumed to be randomly sampled according to the *p.d.f* $\nabla p_f(x)$.

Given arbitrary function $h(x)$, we have the following formula:

$$\int h(x) p_f(x) dx \approx \frac{1}{N} \sum_i^N h(x_i) \qquad (7)$$

So, formula (6) can be reduced into

$$\int W(x,\mu) \| \nabla p_f(x) \|_2 \hat{C}\left(\frac{\nabla p_f(x)}{\| \nabla p_f(x) \|_2} \right) dx \approx \frac{1}{N} \sum_{i=1}^N W(x_i,\mu) \frac{\| \nabla p_f(x) \|_2}{p_f(x)|} \hat{C}\left(\frac{\nabla p_f(x)}{\| \nabla p_f(x) \|_2} \right) \qquad (8)$$

This is a summation weighted by the inverse of the probability $p_f(x_i)$. The formula (8) suppresses the effect of the local descriptors of high probability.

Then, we induce the normalized gradient in as (8):

$$\frac{\nabla p_f(x)}{p_f(x)} \approx \frac{\nabla p_f(x)}{p_g(x)} = \frac{\sum_{i=1}^N x_i g(\| x - x_i \|_2^2)}{\sum_{i=1}^N g(\| x - x_i \|_2^2)} - x \overset{\Delta}{=} \hat{\nabla} p_f(x) \qquad (9)$$

Where the profile g is approximately applied to the normalization on account of $p_f(x) \approx p_g(x)$.

By introducing the normalized gradient (9) into (8), we finally gain the aggregation form to construct features as the histogram of the oriented *p.d.f* gradients. Let $\mu_k, k = 1,...,M$ are the *k-th* word center, and the aggregation around μ_k can be given by:

$$d_k = \frac{1}{N} \sum_{i=1}^N W(x_i,\mu_k) \| \hat{\nabla} p_f(x_i) \|_2 \hat{C}\left(\frac{\hat{\nabla} p_f(xi)}{\| \hat{\nabla} p_f(xi) \|_2} \right) \qquad (10)$$

These features around the respective words are concatenated into the final feature vector:

$$d = [d_1^T, ..., d_M^T]^T \in \square_+^{2dM}. \qquad (11)$$

3.4 *k*-NN Re-Ranking

On the basis of the above, we can further use the top-*k* retrieved image to refine our retrieval results.

Considering a query image, the rank of a database image according to *S** is denoted by $R(Q,D)$. Let N_i be the query's *i-th* retrieved image. Obviously $R(Q,N_i)=i$.

Accordingly $q=\{Ni\}_{\{i=1,\cdots,k\}}$ is the query's *k-NN* re-ranking, as shown in Figure 1**.**

In most cases, the majority of these *k*-nearest neighbors include the same object as in the query image. For example, Some images with the same object are not visually close to the query, and are ranked very low, owning to the features are variant to view point change, object deformation or occlusion,

We also use each localized object in N_q as a query and perform retrieve. The rank of a database image D when using N_i as the query is $R(N_i;D)$, as shown in Figure 1.

According to the rank, we allocate a score $1=R(N_i;D)$ to each database image. The weighted total scores of the database images are then collaboratively determined as:

$$\bar{S}(Q,D)=\frac{\omega_O}{R(Q,D)}+\sum_{i=1}^{k}\frac{\omega_i}{R(Q,D)} \tag{12}$$

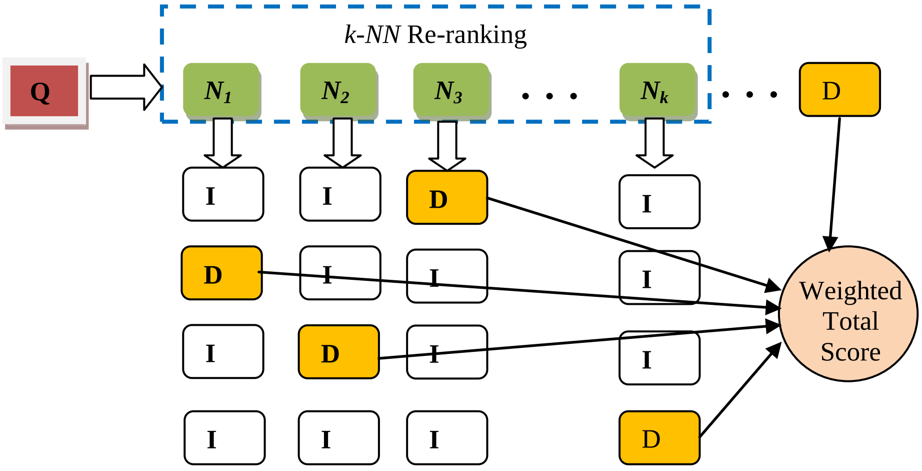

Figure 1. Illustration of *k-NN* Re-Ranking

Rank of N_i in the first retrieval. We set $w_0=1$ and $\omega_i=\frac{1}{R(Q,N_i)+1}=\frac{1}{i+1}$. Query itself can be regarded as the *0-th* nearest neighbor, and formula (12) is accordingly rewritten as:

$$\bar{S}(Q,D)=\sum_{i=0}^{k}\frac{\omega_i}{R(N_i,D)}=\sum_{i=0}^{k}\frac{1}{(i+1)R(N_i,D)} \tag{13}$$

We also consider the rank of the query in each of its nearest neighbors' retrieval results, i.e. $R(N_i;Q)$. Query Q and its nearest neighbor N_i are close only if $R(Q;N_i)$ and $R(N_i;Q)$ are both high.

We revise the weight w_j to be

$$\omega_i=\frac{1}{R(Q,N_i)+R(N_i,Q)+1}=\frac{1}{i+R(N_i,Q)+1} \tag{14}$$

The weighted total scores of database images are determined by:

$$\bar{S}(Q,D)=\sum_{i=0}^{k}\frac{1}{(i+R(N_i,D)+1)R(N_i,D)} \tag{15}$$

Images are then re-ranked based on $\bar{S}(Q,D)$.

Generally in most cases, the first iteration brings significant performance improvement. Furthermore; we may use the new top-*k* retrieved images to perform re-ranking iteratively.

Owning to the *k-NN* re-ranking method can ignore those irrelevant features. The presented method takes advantage of the localized objects in the retrieved images. Moreover, our re-ranking method is no sensitive to false retrieval results in N_q.

In our experiments, the score is related to the ranking. An image will not be re-ranked very highly unless it is close to the query and the majority of those *k-NN* images. However, the weight corresponding to this outlier is relatively small as the rank itself in the query's retrieval list is not high.

On the contrary, a relevant image is close to several images in N_q and will have a high score. Experimental results show our method is robust to the selection of number *k*, even if *k* is large and there are many outliers in N_q, the retrieval accuracy can maintain very high.

Since our method is not sensitive to outliers, no spatial verification is needed. Also, re-ranking can be efficiently performed on the various datasets.

4. Experiments and Analysis

We describe the datasets and features which used in our experiments. *PASCAL-VOC 2007* dataset [1] is used to analyze the performances of the proposed method.

PASCAL-VOC 2007 dataset has *5,011* training images and *4,952* test images. The dataset includes objects of *20* categories and it poses a challenging task of image retrieval due to significant variations in accordance with appearances and poses even with occlusions.

4.1 Results of *k-NN* Re-Ranking

The performance is evaluated by the standard *PASCAL* protocol which computes *mAP* (mean average precision) based on the precision curve. The following five topics are possible in the presented method.

Table 1. Performance Analysis on *PASCAL-VOC 2007*

(a) Parameter *h*

h=1	0.2	0.1	0.05
0.6052	0.6061	0.6173	0.5956

(b) Orientation coding

Random bases	PCA bases
0.6013	0.6173

(c) Component weighting

None	Inverse eigenvalues
0.6021	0.6173

(d) *p.d.f* gradient

None	Normalized
0.5986	0.6173

(1) Parameter *h*.

We utilize the function $f(z)=e^{-z/2h}$ with $h = 0.1$. Because of the curse of dimensionality, such adaptive parameter selection becomes less effective in the higher-dimensional space, since the data samples are sparsely distributed around each sample point.

We apply four $h \in \{1, 0.2, 0.1, 0.05\}$ to the function which are superimposed over the distribution of the distances. The profile of *h=0.1* appropriately obtains the neighboring samples, while those of the other cover too small or too large portion of neighbors.

The favorable performance is acquired at *h=0.1* as shown in Table 1(a); the larger *h = 1* causes better performance than the smaller one *h = 0.05*, showing that it is favorable to gain somewhat large amount of neighbors for constructing discriminative *p.d.f* gradients. In our experiment, we employ *h=0.1, k=25*, and *512* features.

(2) Orientation coding

We focus on the way of coding *p.d.f* gradient orientations. Those orientations are coded using the *PCA* basis vectors. For the alternative to the *PCA* bases, we can also adopt the random orthonormal bases to code them.

The performance comparison is shown in Table 1(b), demonstrating that the *PCA* bases substantially improve the performance. In such a case of complete set of orientation bases, which is smaller than over complete one, the data-driven bases provided by *PCA* effectively code the orientations. Similar, we employ *h=0.1, k=25*, and *512* features.

(3) Component weighting.

The effectiveness of the weighting by the inverse of the *PCA* eigenvalues is shown in Table 1(c) with comparison to the case without weighting. The performance is improved by the weighting which suppresses the orientations commonly occurring across the categories while enhancing the less-frequent but discriminative ones.

In case that we simply use original *p.d.f* gradient ∇p_f without normalization in formula (10), the performance is deteriorated as shown in Table 1(d).

The method employing ∇p_f amounts to the aggregation weighted by the probability p_f which would highly enhance the samples frequently found in the image. In addition, the mean-shift vector $\hat{\nabla} p_f$ is stable in that it always points to the direction where the *p.d.f* is increased. We use the same parameters which described above.

(4) Number of features.

We exhibit the performances on various numbers of features $M \in$ {*64, 128, 256, 512, 1024*} in Table 2. According to the Table 2, we can know that the performance of the proposed method in *PASCAL-VOC 2007*. The proposed method effectively works for object recognition. We employ *h=0.1*, and *k=25.*

Table 2. Performances on Various Numbers of Bag-of-words

	64	128	256	512	1024
BOW+Re-ranking	0.5261	0.5572	0.5822	0.6173	0.6243

The proposed method obtains high performances even on the small amount of features. These results demonstrate that the *p.d.f* gradient orientations more effectively characterize the distribution of the local descriptors.

Since the performance is sufficiently improved by *256* and *512* features, in the following experiments, we apply the proposed method with *256* and *512* features.

In the implementation of the voting-based method, we switch off rotation in *PASCAL-VOC 2007* as most of these query objects are upright.

(5) Number of nearest neighbors *k*

Table 3 shows the performance on *PASCAL-VOC 2007* when we change *k*. Even with only *25* nearest neighbors, the *mAP* is already improved to *0.6173*. When the *k-NN* set N_q becomes larger, the *mAP* keeps increasing.

Table 3. Performances on Various Numbers of Nearest Neighbors *k*

	5	10	15	20	25	35
BOW+Re-ranking	0.5902	0.5979	0.6011	0.6127	0.6173	0.6218

Similarly, since the performance is sufficiently improved by *25*-Nearest Neighbors, we apply the proposed method with *25*-Nearest Neighbors in the following experiments.

4.2 Comparisons to Other Methods

(1) Comparisons to baseline bag-of-words model

We compared *BOW and BOW+Re-ranking* with the baseline bag-of-words model. The results are shown in fig 2. We can see our method significantly outperforms the bag-of-words model. Moreover, the *mAP* of the baseline method increases from *0.4932* to *0.5026*, *0.5318* and *0.5472* respectively, while our method is increases from *0.5602* to *0.5845*, *0.6027* and *0.6173* respectively. This manifests our method is more scalable to larger databases.

(2) Comparisons to *LSH*

We compared *BOW and BOW+Re-ranking* with *LSH also*. In each case, the optimum choice of parameters that maximizes the speedup for a given precision is used.

It can be seen that the *k-NN* re-ranking method performs better than the *LSH* in fig 2, while for the *PASCAL-VOC 2007* dataset our method is faster than *LSH* for all precisions. This also shows that the algorithm proposed scales well with respect to the dataset size.

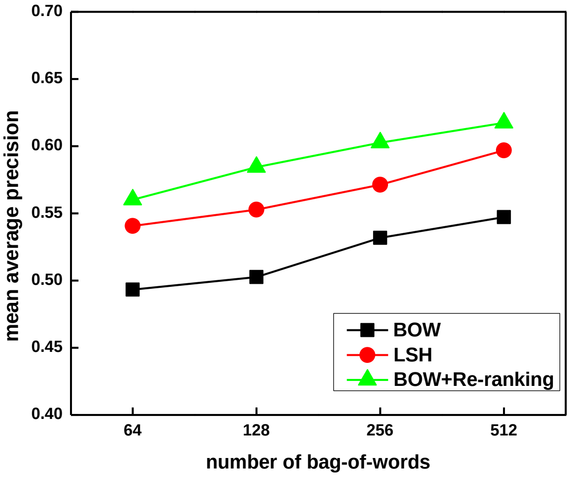

Figure 2. Mean Average Precision Using Various Numbers of Bag-of-words.

Note is that the *LSH* implementation requires significantly more memory compared to *LSH* method for when high precision is required. Although, there are many irrelevant images in N_q when k is large, our approach can still achieve very high accuracy in that case, which demonstrates the robustness of this rank, based method to outliers.

Figure 2 shows the performance of *k-NN* re-ranking. It further significantly improves the retrieval performance, indicating that our method is robust to distractors.

Meanwhile, Figure 2 shows the comparisons of our method with other methods using different number of features. The results of our method are among the best on *PASCAL-VOC 2007*.

4.3 Scalability for Large Datasets

For comparison, we then apply the proposed method to the datasets of *INRIA Holidays*

[20], *MIT-Scene* [21] for scene retrieval and [4] for object retrieval.

INRIA Holidays contains *1,491* images of *500* scenes and objects. One image per scene is used as query to retrieve within the remaining *1,490* images and accuracy is measured as the *mAP* averaged over the *500* queries.

MIT-Scene contains *15,620* images from *67* indoor scene categories and all images have a minimum resolution of *200* pixels. This retrieval task is very challenging due to the large within-class variability and small between-class variability in a large number of categories.

See the Figure 3 and Figure 4, the proposed method exhibits the favorable performance compared to the others, though the improvement is not so significant. However, it should be noted that the dimensionality of the proposed feature with *256* and *512* features to improve the retrieval performance.

Caltech-256 contains *256* object categories and *30,607* images besides a background category in which none of the images belonging to those *256* categories. The intraclass variances regarding such as object locations, sizes and poses in the images are quite large.

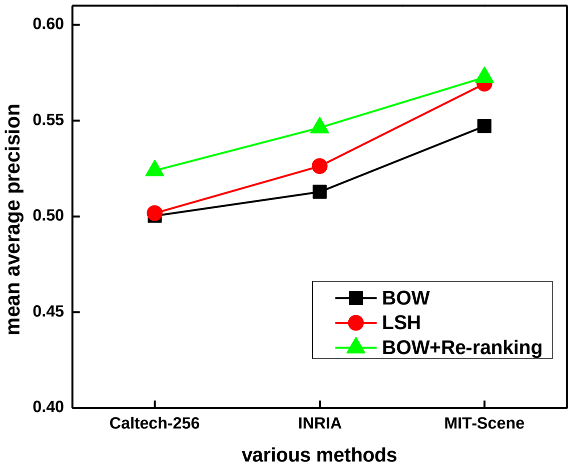

Figure 3. Mean Average Precision while *k*=25, Number of Bag-of-words=256

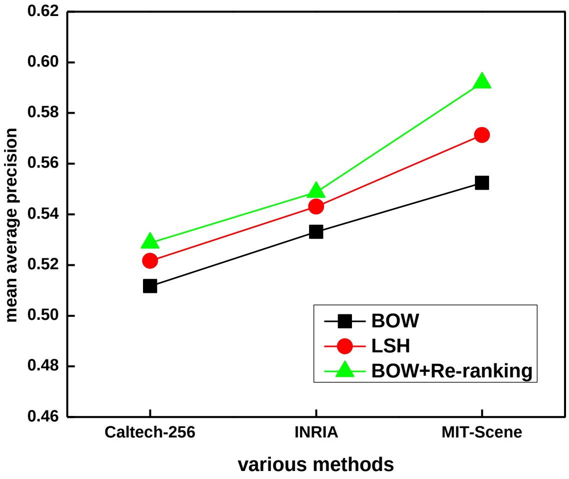

Figure 4. Mean Average Precision while *k*=25, Number of Bag-of-words=512

We report the averaged retrieval accuracy over two methods, and the results are shown in Figure 3 and Figure 4.

It can be seen that the retrieve performance scales well with the dataset size and it benefits considerably from using multiple parallel processes.

In addition to the improved retrieve performance; using multiple parallel processes on a compute cluster has the additional benefit that the size of the dataset is not limited by the memory available on a single machine.

5. Conclusions

We achieve simultaneous image retrieval by utilizing a new BOW method. The proposed method is built upon the probability density function acquired by applying kernel density estimator to those local descriptors. The method exploits the oriented p.d.f gradients to effectively characterize the p.d.f. The proposed method extracted image features, and thus it is applicable to any kinds of image retrieval tasks. Meanwhile, a *k*-NN re-ranking method is further proposed to improve the retrieval performance.

These experimental results exhibit that the proposed method over other methods on *PASCAL-VOC 2007*, *MIT-Scene*, *INRIA Holidays* and *Caltech-256* datasets. It should be attended again that the parameter setting, especially the h = 0.1, is shown to be robust, while the performances might be further increased by tuning the parameter setting carefully in each dataset.

Extensive estimation on several datasets demonstrates that our method increases the performances more significantly on the more challenging datasets due to its high discriminative power. Our method can be integrated in a classification or retrieval system with other components to better image classification or retrieval performance.

Acknowledgments

The authors would like to thank the reviewers for their critical and constructive comments and suggestions.

This paper is supported by the National Natural Science Foundation of China (Grant No.61402420), and the Science and Technology Foundation of Henan Educational Committee (No.14B520049).

References

[1] G. Csurka, C. Bray, C. Dance, L. Fan. Visual categorization with bags of keypoints. In ECCV Workshop on Statistical Learning in Computer Vision, (2004), pp: 1-22.

[2] N. Dalal, B. Triggs. Histograms of oriented gradients for human detection. In CVPR, (2005), pp: 886-893.

[3] The PASCAL Visual Object Classes Challenge 2007 (VOC2007). http://www.pascal-network.org/challenges/VOC/voc2007/index.html.

[4] G. Griffin, A. Holub, P. Perona. Caltech-256 object category dataset. Technical Report 7694, Caltech, (2007).

[5] H. J′egou, M. Douse, C. Schmid, P. P′erez. Aggregating local descriptors into a compact image representation. In CVPR, (2010), pp: 3304-3311.

[6] D. Comaniciu, P. Meer. Mean shift: A robust approach toward feature space analysis. IEEE Transaction on Pattern Analysis and Machine Intelligence, 24(5), (2002), pp: 603-619.

[7] L. Bo, X. Ren, D. Fox. Hierarchical matching pursuit for image classification: Architecture and fast algorithms. In NIPS, (2011), pp: 2115-2123.

[8] J. Philbin, O. Chum, M. Isard, J. Sivic, A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In CVPR, (2008).

[9] A. Mikul´ık, M. Perd'och, O. Chum, J. Matas. Learning a fine vocabulary. In ECCV, 6313, (2010), pp:1-14.

[10] H. J´egou, H. Harzallah, C. Schmid. A contextual dissimilarity measure for accurate and efficient image search. In CVPR, (2007), pp:1-11.

[11] S. Zhang, M. Yang, X. Wang, Y. Lin, Q. Tian. Semantic-aware co-indexing for near-duplicate image retrieval. In ICCV, (2013).

[12] L. Zheng, S. Wang. Visual phraselet: Refining spatial constraints for large scale image search. Signal Processing Letters, IEEE, 20(4), (2013), pp: 391-394.

[13] Y. Xia, K. He, F. Wen, and J. Sun. Joint inverted index. In ICCV, (2013), pp:1-8.

[14] O. Chum, A. Mikul´ık, M. Perd'och, J. Matas. Total recall II: Query expansion revisited. In CVPR, (2011).
[15] G. Tolias, Y. Avrithis. Speeded-up, relaxed spatial matching. In ICCV, (2011).
[16] D. C. G. Pedronette and R. da S. Torres. Exploiting contextual spaces for image re-ranking and rank aggregation. In ICMR, (2011).
[17] M. Charikar. Similarity estimation techniques from rounding algorithms. In ACM STOC, (2002). 2.
[18] Wang J, Kumar S, Chang SF. Semi-supervised hashing for large scale search. IEEE Trans Pattern Anal Mach Intell. (2012) Dec; 34(12):2393-406.
[19] Y. Gong, S. Lazebnik. Iterative quantization: A procrustean approach to learning binary codes. In CVPR, (2011), pp: 817-824.
[20] H. J´egou, M. Douze, C. Schmid. Hamming Embedding and Weak Geometric Consistency for Large Scale Image Search. In ECCV, (2008), 5302, pp:304-317.
[21] A. Quattoni, A. Torralba. Recognizing indoor scenes. In CVPR, (2009), pp:413-420.

Authors

Pang Haibo, he was born in 1979, PH.D, and Lecturer. His research interests include image processing and pattern recognition.

Liu Chengming, he was born in 1979, PH.D, and Lecturer. His research interests include image processing.

Zhao Zhe, she was born in 1983, Master, and Lecturer. Her research interests include image processing.

Li Zhanbo, he was born in 1965, Master, and Professor. His research interests include image retrieval and network security.