Improve VLAD Using the Entropy Produced by BOW

Hongwei Zhao[a], Yeran Wang[b], Pingping Liu [c], Chaoran Zhao[d] and Xiang Li[e]

School of Jilin, Jilin University, Changchun 130012, China;

[a]zhaohw@jlu.edu.cn, [b]wangyr15@mails.jlu.edu.cn, [c]liupp@jlu.edu.cn, [d]zhaocr15@mails.jlu.edu.cn, [e]13051744716@163com

Keywords: Image retrieval, VLAD, BOW, entropy.

Abstract. VLAD (vector of locally aggregated descriptors) is a type of global features extracted from the image, which is always used in image retrieval. Although VLAD is effective, it still needs to improve. VLAD is obtained by accumulating residuals, ignoring the number information of descriptors in a cluster. BOW (bag of words) is also a type of features, which describe the amounts of the descriptors within a cluster. In this paper, the concept of the quantity entropy is proposed based on the BOW method. The amount of descriptors is processed by the calculation steps of entropy to obtain the quantity entropy, and we add the quantity entropy to VLAD, which makes the image retrieval ability of VLAD improved.

1. Introduction

Image retrieval is popular in recent years. We use the features extracted from an image to match with other images to find the similar ones, which implements the image retrieval. Image features includes local features and global features. Many local features have been proposed, like SIFT (scale-invariant feature transform) [1], SURF (speed up robust feature) [2] and so on. And SIFT is one of the most widely used local features. BOW feature [3] and VLAD feature [4] are the global features obtained using SIFT and cluster centers. For BOW method, it trains the cluster centers offline firstly. Secondly, SIFT descriptors of an image are assigned to the cluster centers. And the amount of SIFT descriptors within a cluster is used to establish the histogram, which presents BOW feature. BOW feature contains the amount information of SIFT descriptors [5]. VLAD method is improved on the basis of BOW. Firstly, VLAD also assigns SIFT descriptors to the corresponding cluster center. Secondly, the descriptors in a cluster calculate with the cluster center to obtain residuals. Then the residuals within a cluster are accumulated to obtain the VLAD feature. VLAD feature contains data information of SIFT descriptors [6].

In this paper, we try to describe the image with both data information of descriptors and amount information of descriptors. So we transform the BOW feature by the calculating process of entropy [7], and the transformed BOW histogram is named as the quantity entropy. The quantity entropy is combined with the VLAD as a new feature to represent the image, which we named BVLAD, and the new feature can improve the accuracy of image retrieval.

2. Entropy review

The entropy concept is proposed by T. Clausius in 1854, and it is used in thermodynamics field firstly. Entropy is a type of statistical tools to describe a system.

Assuming x_i is a variable in a system, and the amount of x_i is k, so that $i = 1,...,k$. $p(x_i)$ is the probability when the variable is x_i. The formula to calculate entropy is shown in (1).

$$H(x) = -\Sigma\, p(x_i)\log p(x_i). \qquad (1)$$

3. Improve VLAD using the quantity entropy

There are four steps to produce the new feature. Firstly, we train cluster centers and SIFT descriptors, and produce VLAD in this step. Secondly, we produce the BOW feature for the image.

Thirdly, we use the calculating process of entropy to transform the BOW feature to obtain the quantity entropy. Finally, we combine the quantity entropy and VLAD feature as a new image feature. We describe our method below in details.

3.1 Produce VLAD

VLAD is the global feature produced by cluster centers and SIFT descriptors. Firstly, the cluster centers need to be trained offline by K-means [8] method. The number of the cluster centers is defined as *k*. The dimension of SIFT descriptors is defined as *d*. So that the size of the cluster centers is *k*×*d*. Secondly, SIFT descriptors are assigned to the related cluster center by the nearest neighbor algorithm [9]. Thirdly, we do subtraction between SIFT descriptors within a cluster and the cluster center to produce residuals. Finally, we add the residuals in a cluster together, and we can obtain a vector with the size of *k*×*d*, which is VLAD feature for this image.

3.2 Produce BOW

We count SIFT descriptors in each cluster, and use amounts of descriptors within a cluster build histogram. Because the number of the cluster centers is *k*, the histogram has *k* bins, and a bin stores the amount of SIFT descriptors in a cluster (*i* is the number for each bin, and it is also the number for each cluster). BOW feature stores the amount information of SIFT descriptors, and the size of it is *k*×1.

3.3 Produce the quantity entropy

Because the amounts of SIFT descriptors are almost different for different images, the amount information need to process for its comparable ability. We use the BOW histogram stores SIFT descriptors' amounts for each cluster. We normalize each bin in the histogram using the amount of all SIFT. And the *i*th bin in the normalized histogram represents the probability that the descriptor appears in this cluster. These probability values can be calculated using formula (1) to get the quantity entropy. Through formula (1), which can be seen as a data processing method like normalization, these probability values are limited to a certain range, which makes the BOW histogram more suitable to compare. The quantity entropy has the same size as BOW feature, which is *k*×1.

3.4 Obtain BVLAD

We combine VLAD feature and the quantity entropy as a new feature, which is named as BVLAD. The size of VLAD is *k*×*d*, and the size of the quantity entropy is *k*×1, so the size of BVLAD is *k*× (*d*+1).

4. Experiments

4.1 Image Datasets

There are two image sets we experiment our method in this paper. The first set is INRIA Holidays dataset [10]. The Holidays dataset contains some personal holiday's photos. The set of images are widely accepted to test the robustness for rotations, viewpoint and illumination changes, etc. The Holidays dataset contains a very large variety of scene types, and the scenes includes natural, man-made, water and fire effects, etc. There are 500 image groups in this dataset, and each group represents a distinct scene or object. The second set is Oxford dataset [11]. The Oxford dataset is consisted by 5062 images, so that we also call this dataset as Oxford5k. And these images are collected from Flickr using particular Oxford landmarks. The Oxford5k has 11 groups for different landmarks. This dataset provides 55 queries to evaluate an object retrieval system.

4.2 Results Analysis

We experiment our method on the two datasets above. We train five types of cluster centers, and the amounts of the centers are 16, 32, 64, 128 and 256. We use four types of methods. The first method is the original VLAD, and the results produced by this method are as the baseline values in this paper. The second method is adding the quantity entropy to VLAD, and we mark this method as Method1. We use powernorm [12] in the third method. Powernorm is a type of normalization method, which can improve VLAD feature. For the forth method, we add the quantity entropy to the third method, which we call Method2.

The mAP (mean average precision) value is the estimated standard for Holidays dataset and Oxford5k dataset, and it takes the ranking of the search results into account.

The image retrieval results of Holidays dataset are shown on Table1. When k = 16, the mAP produced by VLAD is 0.489. When experimented by Method1, the mAP of VLAD can be improved to 0.493. Using powernorm method makes mAP value improve to 0.508. Method2 improves the result on the basis of powernorm, and the mAP is 0.513. With the increasing of the number of cluster centers, the mAP values are rising for each method. So the best values for each column are shown in the last row, where k = 256. And the best result on Holidays is 0.604 produced by Method2.

Table 1. The mAP results on Holidays

k	VLAD	Method1	powernorm	Method2
16	0.489	0.493	0.508	0.513
32	0.501	0.509	0.525	0.534
64	0.521	0.532	0.556	0.558
128	0.552	0.557	0.578	0.581
256	0.575	0.577	0.598	0.604

The image retrieval results of Oxford5k dataset are shown on Table2. All of the mAP values are worse than them on Holidays. The results of Method1 and powernorm method are both better than the original VLAD method on the same row, and the best results are produced by Method2. When k = 256, the mAP values are best for each method. And the best mAP on Oxford5k dataset is 0.471.

Table 2 the mAP results on Oxford5k

k	VLAD	Method1	powernorm	Method2
16	0.324	0.326	0.325	0.327
32	0.350	0.354	0.352	0.358
64	0.393	0.395	0.394	0.396
128	0.421	0.423	0.434	0.437
256	0.455	0.459	0.465	0.471

The mAP values produced on the two public datasets show that our method can improve the image retrieval ability of VLAD feature. And when our method combined with powernorm method, the results can be improved further more.

5. Summary

Most of research scholars try to improve VLAD through some complicated ways. We think that there are many useful information in SIFT descriptors themselves. In this paper, we add the amount information of SIFT descriptors to VLAD feature. In order to make the the amount information more comparable, we use BOW feature store it, and process BOW feature by the calculation steps of entropy. The processed information is called the quantity entropy. We can get better retrieval results using the new VLAD feature, which is added the quantity entropy.

This work is supported by the State Key Laboratory of applied optics, the National Natural Science Foundation of China under Grant 61101155, the Jilin Province Science and Technology Development Program under Grant 20150520063JH, and the General Financial Grant from China Postdoctoral Science Foundation under Grants 2015M571363.

References

[1]. LOWE D G. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision. Vol. 60 (2004) No. 11, p. 91-110.

[2]. Yang Z, Shen D, Yap P T. Image mosaicking using SURF features of line segments. Plos One. Vol. 12 (2017) No. 3, p. e0173627.

[3]. KIM T E, KIM M H. Improving the search accuracy of the VLAD through weighted aggregation of local descriptors ☆. Journal of Visual Communication & Image Representation. Vol. 31 (2015) No. C, p. 237-252.

[4]. JEGOU H, PERRONNIN F, DOUZE M, et al. Aggregating Local Image Descriptors into Compact Codes. Pattern Analysis & Machine Intelligence IEEE Transactions on. Vol. 34 (2012) No. 9, p. 1704-1716.

[5]. J GOU H, DOUZE M, SCHMID C, et al. Aggregating local descriptors into a compact image representation. IEEE Xplore Computer Vision and Pattern Recognition. Vol. 238 (2010) No. 6, p. 3304-3311.

[6]. LIU Z, WANG S, TIAN Q. Fine-residual VLAD for image retrieval. Neurocomputing. Vol. 173 (2016) No. P3, p. 1183-1191.

[7]. R NYI A. On measures of entropy and information. Procfourth Berkeley Sympon Mathstatist & Probunivof Calif. Vol. 1 (2015) No. 5073, p. 547-561.

[8]. Wang G, Yao J, School B. Forecasting Model of Fuzzy Time Series Based on Kmeans Algorithm. Acta Analysis Functionalis Applicata. Vol. 31 (2015) No. 9, p. 2103-2116.

[9]. Keller J M, Gray M R, Givens J A. A fuzzy K-nearest neighbor algorithm. IEEE Transactions on Systems Man & Cybernetics. Vol. SMC-15 (2012) No. 2, p. 580-585.

[10]. LIU Z, LI H, ZHOU W, et al. Making Residual Vector Distribution Uniform for Distinctive Image Representation. IEEE Transactions on Circuits & Systems for Video Technology. Vol. 26 (2016) No. 2, p. 375-384.

[11]. Philbin J, Sivic J, Zisserman A. Geometric Latent Dirichlet Allocation on a Matching Graph for Large-scale Image Datasets. International Journal of Computer Vision. Vol. 95 (2011) No. 2, p. 138-153.

[12]. ZHOU Q, WANG C, LIU P, et al. Distribution Entropy Boosted VLAD for Image Retrieval. Entropy. Vol. 18 (2016) No. 9, p. 311-329.