Application Of Image Retrieval Method Based on BOF

XIAO Xue

Chongqing College of Electronic Engineering, Chongqing, 401331, China

gonghengcqgh77@163.com

Keywords: BOF; Image Retrieval; Image Feature

Abstract. For the massive image data and the problem of “dimension disaster”, this paper proposes the BOF-based image retrieval improved algorithm, and combines with VLAD and soft assignment it generates the soft assignment local aggregation descriptor (SA-VLAD) which has a better ability to resist the dimension reduction and a higher recognition rate. When the index mechanism IVFADC is at query time, to ensure the recall ratio and precision rate of the result, the candidates inverted index chain are increased, which leads to the problems of distance calculation and the query time’s increasing. For this point, in the index phase, the scattered distribution is carried out aiming at the database vector, which reduces the burden of distance calculation, and improves the quality of the query results at the same time.

Introduction

With the rapid developments of the Internet and the multimedia technology, people can access to the multimedia data, especially the image data, and the number of the image data increases sharply. How to realize the similarity retrieval of the image content becomes a very important research topic. At present, the domestic and foreign research institutions have achieved many related results, and developed a series of commercial and prototype systems. In domestic, there are TV-FI (TV-Find It) system from Tsing hua University and Web scope CBR from Zhejiang University [1]. Image features. Image features include the high-level semantic feature and the low-level feature. Due to the existence of “semantic gap”, it is difficult for the high-level semantic feature to be obtained directly through the computer, and to retrieve image with the semantic keywords is still a difficult task. And it is another important method of image retrieval to use the similarity matching approach of the image’s low-level feature for image retrieval, but this method depends on the effectiveness of the image feature [2]. For its features of robustness, locality and strong discrimination ability, SIFT feature becomes one of the most commonly used low-level features in the image retrieval field, but its features of high dimension and quantity lead to the difficulty of indexing and querying [3]. BOF (Bag of Feature) method aggregates couple of local feature vectors of an image into a high-dimension vector in the basis of the local feature vectors, through the encoding of the aggregated vector and the organization using the inverted index, which reduces the storage space and the cost of query. However, the price of reducing the feature vectors is that the single feature’s data dimension of the image is very high, which is easy to cause the “dimension disaster”. VLAD (Vectors of Locally Aggregated Descriptors) is the improvement of BOF method, by using the Fisher Kernel thoughts it produces the more simple aggregated descriptors with better query performances. But during the polymerization, VLAD uses the hard allocation strategy to assign the local feature vectors to the clustering categories, which may lead to the situation that two close descriptors be assigned to completely different clustering, thus fails to reflect the distribution features of the vector itself, reducing the efficiency of the retrieval[4,5].

The Generating Algorithm

SA-VLAD uses the k-means method to obtain the k cluster center, and then adopts the distribution principle with the minimum error, uses the soft assignment approach of the fuzzy clustering method for reference to distribute the local feature vectors, finally obtains the membership weight and calculates the aggregation descriptor. The generation process is shown in figure 1.

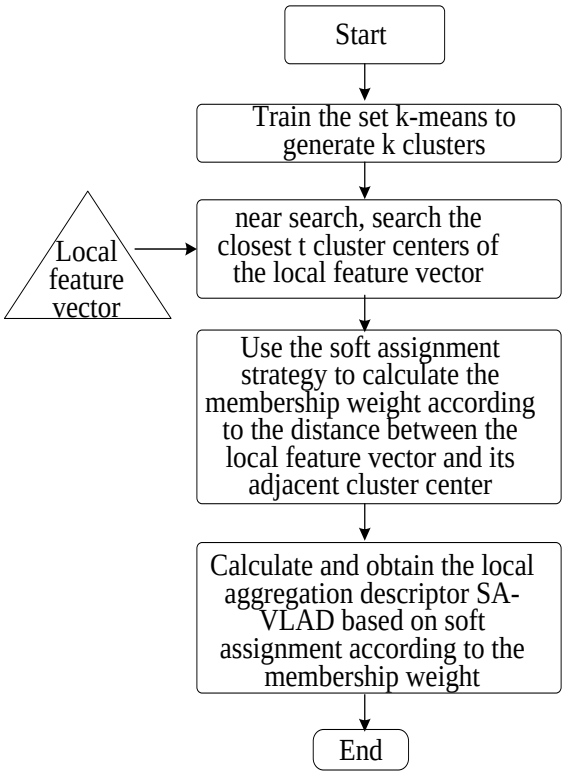

Fig 1. The Generation Schematic Diagram of SA-VLAD

Based on k clusters to carry out the aggregation of the local feature vector to generate the SA-VLAD descriptor, the detailed steps are as follows:

(1) Initialize SA-VLAD to the zero vector s_v with the dimension of $m \times n$. m is the number of the cluster centers and n is the dimension of the image's local feature vector.

(2) Through the near search, the local feature vector S of each image finds out t cluster centers that have the closest distance with the feature vector in the entire cluster center.

$$s(k)_e = s = \left\{ s_i \in codebook \middle| \forall s_j \in codebook / \left(s_i \cup s(t) \right)_m \right\}, \quad \|r - s_i\| \le \|r - s_j\|, 1 \le i \le k, 1 \le j \le k, k \le m \le n \tag{1}$$

In the formula, $s(k)_e$ represents the h close cluster center vector with the vector s. The difference between s and $s(k)_e$ reflects the distribution of s after the mapping of the cluster center, and codebook is the code book vector formed by the aggregation of all the vectors in the cluster center.

Constant t represents the number of the clusters that most feature vectors belong to. When the value of t is too large, the affiliated clusters will be large, the time of clustering will be is long and the clusters with a remote distance from the feature vector will become meaningless for their too small membership weights. If the value of t is too small, it cannot find out the actual subordinate relation between the feature vector and the cluster. Here, the value of t is set to 2.

Calculate the membership weights of feature vector S in the t closest clusters, and use the membership weights to calculate the difference between S and the t closest cluster centers. The collection of the differences between all the local feature vectors in one image and their closest cluster center is namely the SA-VLAD descriptor of the image. sk_i is the vector of d dimension, which stands for the sum of differences of the image's SA-VLAD descriptor in the i cluster center position. The value of sk_i is as follows:

$$sk_i = \sum \frac{1}{t_i(s)} (s - t_i) \tag{2}$$

$$sk = \left(sk_1, ..., sk_1, ...sk_i \right) \tag{3}$$

sk is the SA-VLAD feature value of the image.

SA-VLAD can be directly involved in the sequential retrieval or organize the sequential retrieval by using the index mechanism. However, because SA-VLAD has a high dimension, the dimension reduction is needed before indexing. And it can be known from the characteristics of the image itself that during the category distribution, the image's many local feature vectors, such as SIFT, will be assigned to the same cluster, which means a large number of SIFT vectors have the same adjacent cluster center, and there are clusters without any feature vectors. After the calculation of soft assignment, there may be a large number of consecutive zero values in the cluster center position without any feature points. Thus the linear dimension reduction method can be used to reduce dimension for the high-dimension SA-VLAD. The section 4 of the experiment shows that, to reduce dimension with the same amplitude, SA-VLAD has a smaller loss of precision than VLAD, and the loss of the average precision is even smaller at query time, which states that SA-VLAD has a better ability to resist dimension reduction. After the dimension reduction, the aggregation descriptor can be sequential scanned and queried or be indexed and queried by using any index mechanisms that is suitable for the feature vector.

Experiments

The performance contrast between SA-VLAD and VLAD in the massive database retrieval. Figure 2 shows the recorded values of MAP of VLAD and SA-VLAD under three different index mechanisms in the 1MB image database, which are respectively the values of MAP when there is no index mechanism or under the ADC index mechanism and the DA-IVFADC mechanism after the dimension reduction and the value of T' is 64. In DA-IVFADC, there are $M=16$, $f'=256$, $Lcode=16Bytes$. It can be seen from figure 2 that when the dimension of the aggregated vector reduces from 8192 to 64, the MAP value of SA-VLAD decreases slower than that of VLAD, which states that the former brings out smaller errors than the latter when using PCA to reduce dimension. When the number of data increases, the value of MAP declines, under the same index mechanism, the MAP value of SA-VLAD is better than that of VLAD and the gap is within 0.08, which shows that SA-VLAD still has room for improvement.

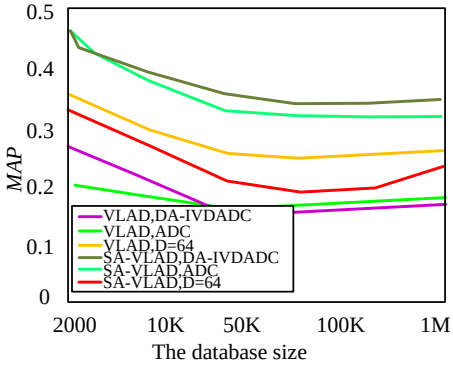

Fig2.The Average Accuracy of SA-VLAD and VLAD under Different Index Mechanisms

The gap of the query time between SA-VLAD and VLAD is time difference for the retrieval object to generate the aggregated vector. By statistics, the average gap of each object's query time is 0.09ms, which shows that when SA-VLAD and VLAD remain the same in not having obvious increasing of query time and storage space, SA-VLAD improves the accuracy of the query results.

And index the SA-VLAD vector in DA-IVFADC when the value of T' is 64, each image only needs to occupy 24 bytes, among which 16 bytes are for codes and 4 bytes are for ID. When the scale of the image library is hundreds of millions, DA-IVFADC+SA-VLAD still can realize the memory-based index in the database.

Figure3 is the typical drawing of the query results combined by VLAD and DA-IVFADC. And figure 14 is the typical drawing of the query results combined by SA-VLAD and DA-IVFADC. In the two figures, the first images in each column are the retrieval images, and each retrieval image returns four query results.

Fig3.The typical drawing of query results combined by VLAD and DA-IVFADC

Conclusion

In this paper the BOF-based image retrieval improved algorithm is proposed, and combines with VLAD and soft assignment it generates the soft assignment local aggregation descriptor (SA-VLAD) which has a better ability to resist the dimension reduction and a higher recognition rate. When the index mechanism IVFADC is at query time, to ensure the recall ratio and precision rate of the result, the candidates inverted index chain are increased, which leads to the problems of distance calculation and the query time's increasing. For this point, in the index phase, the scattered distribution is carried out aiming at the database vector, which reduces the burden of distance calculation, and improves the quality of the query results at the same time.

References

[1]Lowe D.G, Distinctive image features from scale-invariant keypoints, International journal of computer vision, 60(2), 2004, 91-110.

[2]Y. Geng, J. Chen, K. Pahlavan, Motion detection using RF signals for the first responder in emergency operations: A PHASER project[C], 2013 IEEE 24nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), London,Britain Sep. 2013

[3]S. Li, Y. Geng, J. He, K. Pahlavan,Analysis of Three-dimensional Maximum Likelihood Algorithm for Capsule Endoscopy Localization, 2012 5th International Conference on Biomedical Engineering and Informatics (BMEI), Chongqing, China Oct. 2012 (page 721-725)

[4]Y. Geng, J. He, H. Deng and K. Pahlavan, Modeling the Effect of Human Body on TOA Ranging for Indoor Human Tracking with Wrist Mounted Sensor, 16th International Symposium on Wireless Personal Multimedia Communications (WPMC), Atlantic City, NJ, Jun. 2013.

[5]Carson C, Belongie S, Greenspan H, et a]. Blobworid; Image segmentation using expectation-maximization and its application to image querying, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(8), 2002,1026-1038.