Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 14, no. 8, August 2018, pp. 1765-1773
DOI: 10.23940/ijpe.18.08.p13.17651773

Video Retrieval and Sorting Algorithm based on Multiple Features in Sports Videos

Yanbo Su*

Sports Department of Heilongjiang University, Harbin, 150080, China

Abstract

Multimedia information, especially videos, is growing explosively with the rapid development of the Internet and multimedia technology. Due to its variety of image features, it is capable of reaching several hundred dimensions and even thousands of dimensions. Storing and indexing the high-dimensional feature vectors has become key technologies of content-based video retrieval. The residual quantization mechanism, which combines the asymmetric distance and set sorting algorithm based on multi-feature candidates, is improved after analyzing the characteristics of soccer videos. For soccer videos, SD-VLAD (Soft Distribution-Vectors of Locally Aggregated Descriptors), BOC (Bag of Color), and shot type are selected for describing the information of images. To address the problem that the original residual quantized inverted index can only retrieve single features, multiple feature retrieval and sorting are proposed. In the stage of candidate set sorting, a multi-feature based similarity calculation method is designed according to the shots type. The experimental results show that multi-feature hierarchical retrieval and sorting can be achieved at the cost of memory space. While ensuring query speed, the accuracy of the query is improved.

Keywords: sport video; multiple features; contented-based image retrieval; residual quantization

(Submitted on April 25, 2018; Revised on June 13, 2018; Accepted on July 17, 2018)

1. Introduction

The first issue facing large-scale video retrieval is how to extract valid feature information from mass data. The selection and extraction of representative features are directly related to the performance of the content-based video retrieval system. The traditional text-based approach is unsuitable for large-scale video retrieval because of its precision and the need for human intervention. Therefore, the CBVR method is the most mainstream, and it uses videos' low-level visual features and high-level semantic features to describe video information [1].

The global descriptor represents the global feature information of one image. It treats the image as a whole and describes it through a fixed-dimension descriptor vector. Although local descriptors have high robustness for image changes and various occlusion problems, image representations based on local feature vectors usually need to extract a large number of descriptors to fully express the image information. Therefore, the effect of local feature quantity on index efficiency should be considered in large-scale image retrieval. Besides, the quantity of local feature descriptors extracted from each image is not constant, so the index structure based on local descriptors is often complicated. Its advantages include simple expression form, easy retrieval, and less memory space, while its disadvantages include relative sensitivity towards image transformation and occlusion. Furthermore, VLAD (Vectors of Locally Aggregated Descriptors) descriptors [2] and GIST descriptors are commonly used global descriptors [3-5]. The VLAD descriptor refers to the global descriptor that is generated after clustering extracted SIFT descriptors and concatenating the results of each cluster.

The retrieval algorithm is built for the index structure. Its design is directly related to the efficiency of retrieval [6]. The retrieval algorithm mainly involves two kinds of similarity retrieval, scope retrieval and k neighbor retrieval. In the content-based image retrieval system, the characteristic vector and the selection of measuring distance are subjective or tentative; the characteristic vector itself is only an approximate representation of multimedia content rather than a precise one, so

* Corresponding author.
E-mail address: suyanbo777@163.com

researchers have proposed approximate similarity queries. Compared with non-similar retrieval, the similarity query method greatly improves the retrieval speed at a small cost of accuracy, so it is widely used in various index structures. Sorting mainly involves completing the processing of candidate sets, matching the query image with the candidate set accurately, and returning to the user's task from the big to the small according to the similarity score. The ranking strategy based on the global descriptor reduces the sorting complexity and improves the accuracy of the sort. Index generation is conducted off-line, which would not have any significant impact on the user's experience. However, the search algorithm and sorting algorithm are most familiar to users; therefore, their quality directly influences the entire performance evaluation of the video retrieval system. The search algorithm is tailored to the index structure. A good index structure can only exert its advantages through the related search algorithm; thus, the development of an exquisite search algorithm and sorting algorithm has become an important part of designing retrieval systems.

2. Video Feature Extraction

2.1. SD-VLAD Descriptor Extraction

The popular local aggregation descriptors are BOF, min BOF, VLAD, and so on. Compared with the BOF descriptors, VLAD descriptors not only contain the number of visual words, but also include the spatial distribution of local feature vectors relative to visual words. Therefore, VLAD descriptors have significantly better performance than BOF descriptors. However, the traditional VLAD employs a hard-to-allocate strategy whereby a local descriptor vector can only be placed in a cluster closest to it, resulting in a loss of VLAD information. SD-VLAD is an improvement of VLAD, which is a local feature aggregate descriptor generated through the combination of the soft distribution [7] idea with VLAD. The soft-to-allocate idea is to allocate local feature vectors to multiple nearest neighbor clusters, which overcomes the effect of the original VLAD hard-to-allocate strategy and improves the VLAD recognition performance.

2.2. BOC Color Descriptor Extraction

The local vector of the SD-VLAD descriptors is the SURF Feature in 2.1. The SURF feature evolves from SIFT features. It inherits the robustness of SIFT. It only uses the grayscale information of the image and ignores the color information, so the resulting SD-VLAD recognition performance is still unsatisfactory. To solve the above problem, the researchers add BOC color descriptors to complement with the SD-VLAD descriptors and jointly represent the content information of the image. BOC descriptors are an improvement on the color histogram. They differ from the original color histogram mainly in two aspects. Firstly, the trained code book is used to replace the original average color partition [8]; secondly, the color histogram is constructed and standardized by adopting the BOW framework [9] and the improved Fisher kernel [10]. Therefore, it shows more robustness, avoiding the negative impact of the most frequently occurring color on the recognition performance.

3. Sport Video Retrieval based on Multi-Feature

3.1. Asymmetric Distance Calculation

Quantify the query vectors and the database vectors, respectively, and then calculate the distances between their approximate vectors; the above distance calculation is called the symmetric distance calculation. However, the asymmetric distance calculation is the distance calculation between the un-quantized query vectors and the quantified database vectors. Experiments have shown that the asymmetric distance calculation can provide higher distance calculation accuracy.

y is the database vector and x is the query vector. The quantized vector x and vector y are represented by $q(x)$ and $q(y)$ respectively. When using the symmetric distance calculation method, the distance between vector x and vector y can be used to approximate the distance between $q(x)$ and $q(y)$. This is shown in Equation (1).

$$d(x, y) \approx d(q(x), q(y)) \tag{1}$$

When using the asymmetric distance calculation, the distance between vector x and vector y can be used to approximate the distance between $q(x)$ and $q(y)$. This is shown in Equation (2).

$$d(x,y) \approx d(x,q(y)) \tag{2}$$

In the residual quantization coding method, assuming residual quantizer includes L quantizers, each quantizer includes k cluster centers. The residual coding sequence of database vector $\overline{y}$ is u_j after encoding by the residual quantizer. This is an approximate representation of vector quantization, shown in Equation (3).

$$\overline{y} = \sum_{i=1}^{L} \overline{y}_1 = \sum_{i=1}^{L} c_{i,u_i} \tag{3}$$

The Euclidean distance between the query vector x and the database vector $\overline{y}$ approximates the Euclidean distance between the query vector x and the database vector y quantization. This is shown in Equation (4).

$$\begin{aligned} &d(x,y)^2 \approx d(x,\overline{y})^2 = \| x-\overline{y} \|^2 = \| x \|^2 + \| \overline{y} \|^2 - 2(x,\overline{y}) = \\ &\| x \|^2 + \| \overline{y} \|^2 - 2(x, \sum_{i=1}^{L} c_{i,u_i}) = \\ &\| x \|^2 + \| \overline{y} \|^2 - 2\sum_{i=1}^{L} (x, c_{i,ui}) \end{aligned} \tag{4}$$

3.2. Residual Quantization Approximate Retrieval

The incomplete searching idea based on residual quantization is the following: divide the database into multiple sections, then scan the vector code of each section, and finally calculate the exact distance. It is known from the forming process of residual code that database vectors can be put into a data structure that is similar to a tree structure to achieve incomplete retrieval, therefore reducing searching time. When database vectors are quantized through a residual coding quantizer, the vector sequence of the approximate representation of y can be obtained. This is shown in Equation (5).

$$y \approx \sum_{i=1}^{L} c_i \tag{5}$$

In the incomplete retrieval, the vector sequence of the approximate representation is divided into two parts. Assume there are L_1 items in the first part and $L_2 = L - L_1$ items in the second part. We use the sequence in the first part as a rough approximation of y . This is shown in Equation (6).

$$y \approx y^{L_1} = \sum_{i=1}^{L_1} c_i \tag{6}$$

The asymmetric distance between query vectors and the rough approximate vectors can be achieved through a lookup table. This is shown in Equation (7).

$$d(x, y^{L_1})^2 = \| x \|^2 + \| y^{L_1} \|^2 - 2\sum_{i=1}^{L_1} (x, c_{i,u_i}) \tag{7}$$

Delete those vectors whose rough distances between query vectors are greater than a certain threshold value in the database. Conduct a more accurate distance calculation on the remaining candidate vectors. The final distance calculation Equation (8) is as follows:

$$\begin{aligned} &d(x,y)^2 \approx d(x,y^{L_1})^2 = \| x \|^2 + \| y^L \|^2 - 2\sum_{i=1}^{L_1} (x, c_{i,u_i}) = \\ &d(x,y^{L_1})^2 = \| y^L \|^2 + \| y^{L_1} \|^2 - 2\sum_{i=L_1-1}^{L} (x, c_{i,u_i}) \end{aligned} \tag{8}$$

3.3. Multi-Feature Retrieval Process

The multi-feature hierarchical index in Figure 1 only quantifies the database vectors and does not conduct residual quantization on query vectors, and the asymmetric distance calculation is applied in calculating the distance. Use the Euclidean distance between the query vectors x and the approximate vectors $\overline{y}$ of database vectors y to look up approximately the distance between query vectors x and database vectors y.

The detailed process of multi-feature retrieval is as follows:

(1) Extract the BOC descriptor of the query image, SD-VLAD descriptor, and shot type.

(2) Generate the lookup table. The purpose of a lookup table is to speed up precise distance calculations. This step mainly generates two lookup tables of the BOC feature index layer and SD-VLAD feature index layer.

(3) According to the BOC feature of the query image, the nearest W_{boc} inverted list is found in the BOC feature index layer. The calculation Equation (9) of rough distance is as follows:

$$d(x_{boc}, y^{L_1^{boc}})^2 = \| x_{box} \|^2 + \| y^{L_1^{boc}} \|^2 - 22\sum_{i=1}^{L_1^{boc}} (x_{boc}, c_{i,u_i}^{boc}) \tag{9}$$

(4) According to the SD-VLAD feature of the query image, the nearest $W_{sd-vlad}$ inverted list is found in the SD-VLAD feature index layer. The calculation Equation (10) of rough distance is as follows:

$$d(x_{sd-vlad}, y^{L_1^{sd-vlad}})^2 = \| x_{sa-vlad} \|^2 + y^{L_1^{sd-vlad}} \|^2 - 2 \sum_{i=1}^{L_1^{sd-vlad}} (x_{sd-vlad}, c_{i,u_i}^{sd-vlad}) \tag{10}$$

(5) Select the corresponding inverted list according to the shot type of the query. If the shots type of the query image is the far shot, the elements in the far side inverted list should be used as the candidate set to calculate the exact distance. Otherwise, the elements in the non-far inverted list will be selected as candidate sets.

4. Candidate Set Sorting based on Multiple Features

Through the query process of the index, we can quickly obtain a set of disordered results with different similarity. However, among the results that users wish to return, the results with high correlation are generally ranked in the first part, while those with low similarity are ranked in the second part. This is exactly the end that the candidate set sequencing aims to achieve. The similarity score based on the local descriptor is generally calculated through voting, while the similarity score based on the global descriptor is generally calculated through the operation between feature vectors, including the cosine, dot product, Euclidean distance of the included angle, and other operations.

4.1. Sorting based on Linear Combination

The idea of linear combination sequencing is to combine the calculated results obtained from different features and then sort them comprehensively. It is the most direct multi-feature sequencing method. In this kind of method, the similarity score that the single feature has on each image is usually obtained first, and then the similarity scores obtained from different features are weighted averagely to determine the final similarity scores of the query image. The final sequencing result can be obtained by ranking its similarity score in descending order. The similarity calculation Equation (11) based on linear combination is as follows:

$$S = w_1 s_1 + \cdots + w_i s_i + w_n \tag{11}$$

From the above similarity calculation equation, it is known that the linear combination coefficient, the similarity between the characteristic weight, and the single feature all affect the final ranking score.

4.2. Multi-Feature Similarity Calculation for Sports Videos

Due to differences in the content of images, the weight of features has a great influence on the sequencing candidate set. In videos of football matches, long shot images are very different from non-long shot images. Basically, there are only two major parts in long shot images: the spectator area and the court area. There are not many differences between spectators in most long shot images; however, players on the field are small and can barely be identified, so color becomes a primary feature that users care about. On the contrary, non-long shot images include middle-view shots, close-up shots, and out-of-field shots. The local features of SURF can perfectly reflect the information of such shot type images.

According to the above analysis, greater weight shall be given to the BOC descriptor in the long shot, while greater weight shall be given to SD-VLAD in the non-long shot. The similarity score calculation Equation (12) is as follows:

$$S = 1/(w_{boc} d_{boc}^2 + w_{sd-vlad} d_{sd-vlad}^2 + 1) \tag{12}$$

According to Equation (10), the greater the weighted distance between the multiple features of the two images, the lower the similarity, and conversely, the greater the similarity.

4.3. Candidate Set Sorting

For the candidate set of index retrieval, the Equation (12) is used to sort the candidate result set. The detailed steps are as follows:

(1) Accurate distance calculation. Use the generated lookup table and the rough distance calculated previously to calculate the exact distance between the query image VLAD descriptor and VLAD descriptor and the corresponding feature of the database image. It is shown in Equation (13).

$$d(x,y)^2 \approx d(x,y^{L_1})^2 = \| y^L \|^2 + \| y^{L_1} \|^2 - 2\sum_{i=L_1-1}^{L} (x, c_{i,u_i}) \tag{13}$$

(2) Calculate and sort the similarity scores based on the distance and shot type. Due to the huge differences between long shot images and non-long shot images, the weight w_{boc} of the BOC descriptor and the weight $w_{sd-vlad}$ of the SD-VLAD descriptor would vary according to different shot types. The calculation Equation (14) of the similarity score is as follows:

$$S(x,y) = 1/(w_{boc} d_{boc}^2 + w_{sd-vlad} d_{sd-vlad}^2 + 1) \tag{14}$$

(3) Sort according to similarity and return the result. The similarity can be characterized by the integrated distance since the similarity score between the query images and the database images is inversely proportional to the featured weighting square distance. The closer the integrated distance, the higher the similarity score between the two images is, and vice versa. The ascending order of the integrated distance is the descending order of the similarity score.

5. Experimental Results and Analysis

5.1. Experimental Environment

The hardware environment is Intel (R) Xeon (R) CPU 8 core, 16G memory. The operating system is Centos 5.6. The programming environment is Matlab 7.10.0.

5.2. Experimental Data

To verify the multi-feature sequencing strategy and the hierarchical index mechanism based on multi-feature, three databases DB1, DB2, and DB3 are prepared for the multi-feature sequencing strategy. Images of the three databases are a small number of soccer videos and images selected from the Souqiu network, among which database DB1 includes 8,862 long shot images and database DB2 includes 8,056 non-long shot images. The database (DB3) is a combined database of images in DB1 and DB2. The database DB3 sample is shown in Figure 1. There are a total of 82 long shot query images and 51 non-long shot query images. This query performance is evaluated by mAP (mean Average Precision). The larger the mAP value, the better the query performance.

Figure 1. Database image sample

For the multi-feature hierarchical indexing mechanism, database DB4 is prepared, which contains three image sets: training sample collection, index data set, and query image collection. The image in database DB4 is a subset of the Souqiu network image library, where the training set size is 100000 images, the index data set is a total of 1201326 images, and the query image is 50.

5.3. Experimental Results

5.3.1. The Influence of Residual Quantized Parameters on the Quantization Error of SD-VLAD and BOC Descriptors

L is the number of sub-quantizers that the residual quantizer contains. $k = \{64,128,256,512\}$ is the number of cluster centers for each sub quantizer. L and k both determine the $L\log_2 k$ vector.

This is shown in Figures 2 and 3. The residual quantization error would decrease as the number of the sub-quantizer and the cluster center of each quantizer increases. Therefore, it can be considered that the number of sun-quantizers L is inversely proportional to the quantization error. It can also be seen from the figure that when the coding length is fixed, the quantization error of the residual coding quantizer with smaller L and larger k is lower than that of the quantizer with larger L and smaller k. Therefore, k has a greater influence on the quantization error than L. In practical applications, though increasing the value of L and k can reduce quantization error, the storage space would also be increased; therefore, a trade-off between space and accuracy is needed. Experiments show that the two descriptors can achieve a good balance under the condition of k=256 and L=8, which greatly compresses the storage space while ensuring the accuracy of the query.

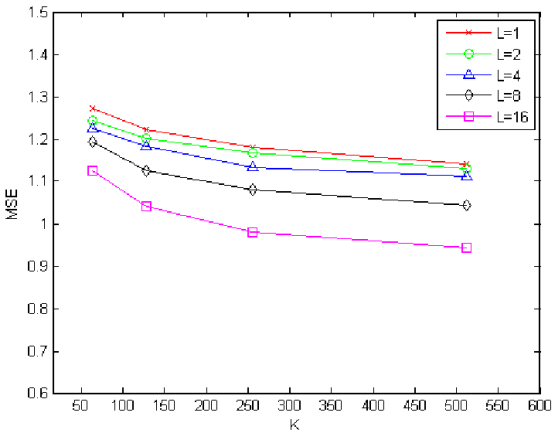

Figure 2. The relation graph of the number of BOC quantization, the number of clustering of the sub quantizer, and the mean square error

5.3.2. Experimental Results of Multi-Feature Sorting Strategy

Figure 4 is a curve diagram of the weight size and the average accuracy endowed to the SD-VLAD descriptor and SD-VLAD descriptor under images of different shot types. $w_{sd-vlad}$ =0 means to only use the SD-VLAD descriptor to indicate the image, while $w_{sd-vlad}$ = 1 means to only use the SDA-VLAD descriptor to indicate the image. In this experiment, the number of cluster centers of SD-VLAD is 16, and the number of color cluster of the SD-VLAD descriptor is 256. Figure 4 shows that the retrieval effect of the SD-VLAD feature obtained by adopting any feature alone is lower than combining the two. Thus, the strategy of combining two features to indicate images improves the accuracy of query results.

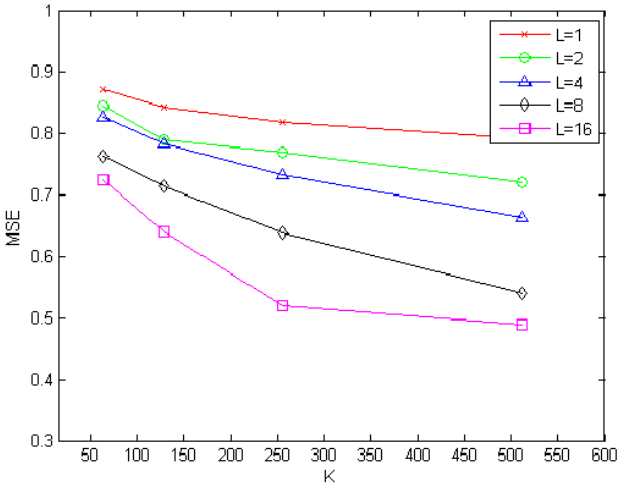

Figure 3. The relation graph of the number of SD-VLAD quantization, the number of clustering of the sub-quantizer, and the mean square error

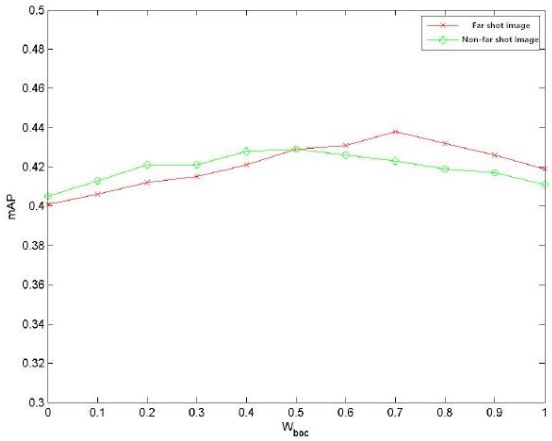

Figure 4. The relationship between characteristic weights of SA-VLAD, SA-VLAD, and mAP for different lens types

It can also be seen from the figure that the optimal weight assignment points of the SD-VLAD descriptor and the SD-VLAD descriptor vary for different shot types. In long shot images, the extracted SIFT local features are limited. Most of the SIFT local features come from the spectator area, which is not representative. The recognition rate of the SD-VLAD descriptors is affected to a certain extent, while the SD-VLAD feature can better reflect the information of images that apply to users' sense organs; thus, the weighting endowed on SD-VLAD descriptors is greater. In non-long shot images, the weight of the two features is relatively balanced because the large number of extracted SIFT features increases the recognition effect of SD-VLAD.

The optimal value of the weight allocation between the SA-VLAD descriptor and the SA-VLAD descriptor needs to be tested and adjusted, and it is closely related to the image of the database. In view of the data set used in this paper, the optimal weight allocation ratio in far shot images is the following: the SD-VLAD descriptor is 0.3, and the SD-VLAD descriptor is 0.7. In non-far shot images, the optimal weight allocation ratio is the following: the SD-VLAD descriptor is 0.45, and the SD-VLAD descriptor is 0.55.

5.3.3. Experimental Results based on Multi-Feature Hierarchical Index

R is the number of returns, and W is the number of columns for the selected inverted list. L_{boc}, k_{boc}, and L_1^{boc} are the number of sub-quantizers of the residual quantizer corresponding to the BOC characteristics, the number of sub-quantizer clustering centers, and the number of sub-quantizers used to establish the index, respectively. $L_{sd-vlad}$, $k_{sd-vlad}$, and $L_1^{sd-vlad}$ are the number of sub-quantizers of the residual quantizer corresponding to the SA-VLAD characteristics, the number of sub-quantizer clustering centers, and the number of sub-quantizer used to establish the index, respectively.

It can be seen from Figure 5 that as *R* increases, the query accuracy of the two indexes declines. However, the accuracy curve of the hierarchical index based on multi-feature is always above the index based on residual quantization, which shows that the query accuracy of the multi-feature hierarchical index is better than the index based on residual quantization for big data football video image libraries.

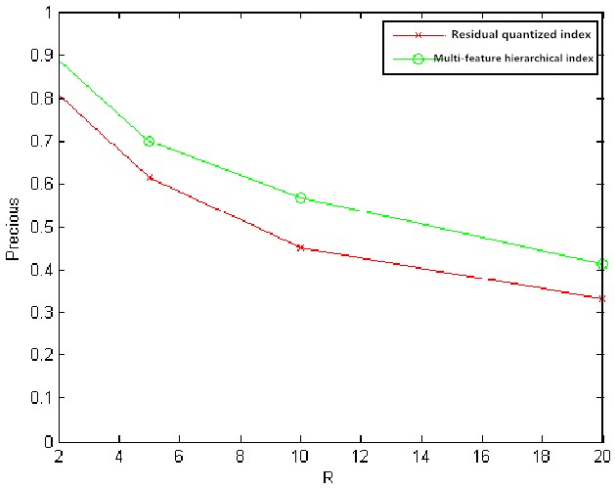

Figure 5. Comparison of query accuracy of different index mechanisms

Figure 6 is the recall rate of the two index mechanisms of the database DB3 under different conditions of *R*. In Figure 6, the query recall rate increases as the query returning results of *R* continuously increase. The recall rate based on the multi-feature index is lower than that of the residual quantization index. The reason behind this phenomenon is that the multi-feature hierarchical index has filtered more data objects during the searching phase of the candidate set. However, the difference between their recall rates is within 0.1.

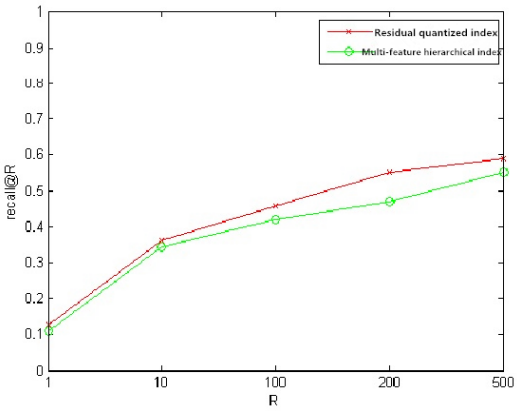

Figure 6. Comparison diagram of query recall rate of different index mechanisms

Table 1 is the comparison of retrieval time between two index mechanisms under different w conditions. The query time in the figure is the sum of three parts: feature extraction time, index retrieval time, and candidate set sequencing time. Because the multi-feature hierarchical index needs to extract various features, it took more time in the feature extraction phase. Each element of the candidate set needs to conduct two accurate distance calculations and asymmetric distance calculations. The approximate neighbor calculation is based on the residual quantization of descriptors at the stage of candidate set sequencing phase. Table 2 is a comparison between the content space and index establishment time cost of two index mechanism indexes.

Table 1. Query time under different w

Indexing mechanism	$w=1$	$w=2$	$w=3$
Residual quantized index	99.7 ms	151.2 ms	442.3 ms
Multi-feature hierarchical index	111.2 ms	165.2 ms	457.2 ms

Table 2. The time and storage space to generate two indexes

Indexing mechanism	Space (G)	Time (min)
Residual quantized index	1.27	400.2
Multi-feature hierarchical index	2.31	754.2

Based on the above experimental results, we can see that the performance of the hierarchical index query based on multi-features is better than that of the inverted index based on residual quantization. However, the larger occupancy of index space requires further research and solution.

6. Conclusion

This paper focuses on the multi-feature hierarchical indexing query process and several technical key points including asymmetrical distance calculation, which is based on the approximate neighbor calculation and residual quantization. This paper proposes a multi-feature similarity calculation method. Experiments show that the multi-feature hierarchical index can improve the accuracy of the query at the cost of memory storage while ensuring query speed. It can provide a better user experience than the single-feature method based on the multi-feature similarity calculation.

Acknowledgement

This work was supported by the Heilongjiang Province University Basic Research Fund (HDJDY201709).

References

1. Y. G. Pu, "A Review of Research on the Key Technology of Content based Video Retrieval," *Information Science*, Vol. 28, No. 3, pp. 464-469, 2010
2. H. Jegou and D. Matthijs, "Aggregating Local Descriptors into a Compact Image Representation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3304-3311, San Francisco, USA, 2010
3. O. Aude and T. Antonio, "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope," *International Journal of Computer Vision*, Vol. 42, No. 3, pp. 145-175, 2011
4. O. Aude, "Building the Gist of a Scene: The Role of Global Image Features in Recognition," in *Proceedings of Progress in Brain Research*, pp. 23-36, 2006
5. M. Douze, "Evaluation of GIST Descriptors for Web-scale Image Search," in *Proceedings of the ACM International Conference on Image and Video Retrieval*, pp. 1-8, 2009
6. H. Jegou and C. Schmid, "Accurate Image Search using the Contextual Dissimilarity Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 1, pp. 2-11, 2010
7. P. James, C. Ondrej, and I. Michael, "Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases," in *Proceedings of the IEEE 9th Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2008
8. J. Michael, "Color Indexing," *International Journal of Computer Vision*, Vol. 7, No. 1, pp. 11-32, 1991
9. S. Josef and Z. Andrew, "Video Google: A Text Retrieval Approach to Object Matching in Videos," in *Proceedings of the IEEE 12th International Conference on Computer Vision*, pp. 1470-1477, 2013
10. P. Florent and M. Thomas, "Improving the Fisher Kernel for Large-scale Image Classification," in *Proceedings of the 11th European Conference on Computer Vision (ECCV)*, pp. 143-156, 2010

Yanbo Su received his B.S degree from Harbin Sport University. He is currently a lecturer at Heilongjiang University. His research interests include physical education, sports training, and sports humanities.