Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 14, no. 5, May 2018, pp. 1069-1077
DOI: 10.23940/ijpe.18.05.p26.10691077

Extraction and Mining of Video Feature in Sport Videos

Yang Han*

Sports Department of Heilongjiang University, Harbin, 150080, China

Abstract

On the basis of analyzing the characteristics of sports video, the parameters of the feature generation are adjusted. According to the sports video library, three features of SD-VLAD (Soft Distribution-Vectors of Locally Aggregated Descriptors), BOC (Bag of Color) and shot type were selected as the description information of the image; the appropriate parameters were selected through experiments; the best parameter configuration for soccer video library was given. In order to detect the influence of parameters in SD-VLAD and BOC descriptors on the recognition effect of descriptors, and select the appropriate parameters, the experiment was carried out in part of the library of search web, and the experimental results were analyzed.

Keywords: sport video; multiple features; bag of color; VLAD

(Submitted on February 1, 2018; Revised on March 19, 2018; Accepted on April 27, 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, CBVR method is the current mainstream, which uses video's low-level visual features and high-level semantic features to describe video information [12].

It is very difficult to achieve retrieval by using low-level visual features to extract high-level semantic features, so currently the video retrieval mainly focuses on the retrieval of low-level features. The existing low-level visual features mainly include color, texture and shape as well as various combinations of features. Low-level features can be divided into local feature and global feature according to different extraction regions. For local features, the feature detection method is first used to detect the feature location or region features, and then the feature extraction method is used to extract local feature descriptors from local images. Usually multiple local feature descriptors can be extracted from one image. Local feature directly determines the quality of the afterward retrieval, classification and identification. Common local descriptors are SIFT [4], GLOH [10], SURF [1], PCA-SIFT [10], etc. The SIFT (Scale-invariant feature transform) is based on the scale space proposed by Lowe, and has a strong robustness image local feature descriptor for images [14].

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 the 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; its disadvantages include relative sensitivity towards image transformation and occlusion. Furthermore, VLAD (Vectors of Locally Aggregated Descriptors) descriptors [7] and GIST descriptors are commonly used global descriptors [2,3,5]. VLAD descriptor refers to global descriptor finally generated after clustering SIFT descriptors extracted from image and concatenating the results of each cluster. VLAD descriptors show good robustness towards occlusion and rotation.

* Corresponding author.
E-mail address: hanyang5722688@126.com

Image feature is an “interesting” part of digital images as well as the starting point for the many computer image analysis algorithms such as image classification and content-based image retrieval. Therefore, the selection of image features often determines whether an algorithm is successful or not. The contents of the image vary widely, yet only by selecting the appropriate image features can the performance of the image retrieval system be improved effectively. A feature can only reflect one aspect of the image information. In order to improve the query precision, it is necessary to combine it with a variety of features. To handle the issue that a single feature cannot fully reflect the image content, a new similarity calculation formula is proposed based on SD-VLAD (Soft Distribution-Vectors of Locally Aggregated Descriptors) descriptors, BOC color descriptors and image lens types, which may improve the precision of retrieval results.

2. SD-VLAD descriptor extraction

With the appearance of local feature descriptors with good recognition performance and the optimization of traditional clustering methods [14], local polymerization descriptors are widely used in image retrieval and classification. The local feature descriptor index is based on the original. Because the storage space is very large, it is difficult to store in memory so most of the indexing mechanism will be part of the data stored in the storage device, the read data I/O cost increase, and slow query. In order to overcome the above problems, local aggregation descriptors emerge.

2.1. SD - VLAD descriptor

The popular local aggregations 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. Due to the variety of local feature sources such as noise in images, non-affine transformations in some image regions and different lighting, local descriptor vectors are not “yes/no” relationships. It is possible that some local descriptors have quite small distance with multiple neighbor clustering centers. If they are blindly allocated to the nearest cluster, a large amount of information will be lost, leading to failure in reflecting the spatial distribution of local features, and eventually resulting in the decreased recognition performance of VLAD descriptors.

SD-VLAD is an improvement of VLAD, which is a local feature aggregate descriptor generated through the combination of the soft distribution [8] 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. The SD-VLAD generation process is shown in Figure 1. SD-VLAD generation process is divided into two parts. First, the k-means method is used to train the training sample sets to generate k clustering center as code books, and each clustering center c_i is a code word. The definition of the encoding codebook Equation (1) is as follows.

$$codebook = \{c_1, ..., c_k\} \tag{1}$$

Second, according to the codebook, the local eigenvector of the aggregated image is a descriptor. Figure 1 is a schematic of the SD-VLAD descriptor generation.

The SD-VLAD descriptor needs to obtain the number of neighbor clustering number t and distance difference threshold α before generation, and then aggregate local characteristics according to the existing clustering center. The specific steps are as follows:

1. Initialize the SD-VLAD vector v and set vector per dimension is set to zero. Dimension is $k*d$. k is the number of clustering center, d is the dimension of local eigenvectors;

2. For the local eigenvector p of each image, in all cluster center, the nearest neighbor t clustering is obtained through the nearest neighbor search. It is shown in Equation (2).

$$c(p)_h = c_i = \begin{cases} c_i \in codebook \mid \forall c_j \in codebook \setminus (c_i \cup c(S)_m) \\ \| p - c_i \| \leq \| p - c_j \| \\ 1 \leq i \leq k, 1 \leq j \leq k, h < m < t \end{cases} \quad (2)$$

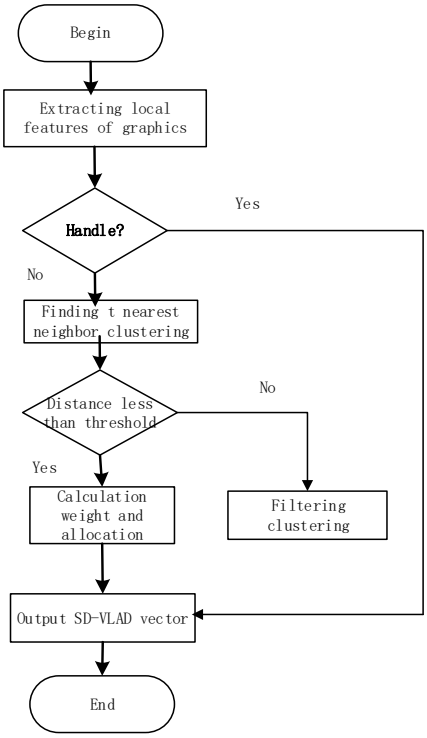

Figure 1. SD-VLAD generating schematic diagram

3. Filter distance too large clustering, $d_i, 1 \leq i \leq t$ is the distance p and its t nearest neighbor cluster centroid. $d_{\min}$ is the distance of its nearest cluster center. p can be assigned to the i nearest neighbor cluster when $d_i - d_{\min} < \alpha$. α is the distance difference threshold. After filtering, there are t' clusters left.

4. According to the distance p to neighbor clustering, weight w_i is assigned respectively. The calculation Equation (3) of w_i is as follows:

$$w_i = \frac{\sum_{h=1}^{t'} (\frac{1}{\| p - c_i}})^2}{(\frac{1}{\| p - c_i}})^2} \quad (3)$$

5. The collection of the residual vectors in each cluster is spliced together to assign weights; SD-VLAD was calculated. v_i is a vector of d dimension, describing the sum of the remaining vectors in the position of the local descriptor at the i th cluster center. According to the distribution weight, v_i is shown in Equation (4) and (5).

$$v_i = \sum w_i (p - c_i) \quad (4)$$

$$v = [v_1, ..., v_k] \tag{5}$$

When the vector v is normalized, it is a SD-VLAD vector. The similarity between the two images can be obtained directly by calculating the Euclidean distance between the SD-VLAD vectors. However, due to the high dimension of SD-VLAD, the CPU computing time is larger in large-scale calculation.

2.2. BOC color descriptor extraction

The local vector of the SD-VLAD descriptors is the SURF Feature in 2.1. SURF feature evolves from SIFT features. It inherits the robustness of SIFT. Also, compared to the SIFT feature, it dramatically improves feature extraction at the expense of a small amount of information loss. However, SURF feature also inherits the shortcomings of SIFT. It only uses the grayscale information of the image and ignores the color information of the image, so the resulting SD-VLAD recognition performance is still unsatisfactory. In order to solve the above problem, the researchers add the 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. It differs from the original color histogram in two main points. Firstly, the trained code book is used to replace the original average color partition [11]; secondly, the color histogram is constructed and standardized by adopting the BOW framework [9] and the improved Fisher kernel [6]. Therefore, it shows more robustness, avoiding the negative impact of the most frequently occurring color on the recognition performance. The generation of BOC descriptors is mainly divided into two steps: training the code book and generating the descriptors. It is shown in Figure 2 and 3.

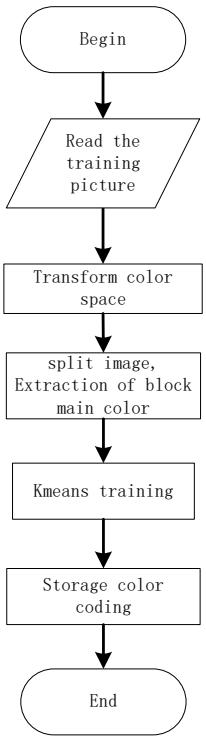

Figure 2. BOC descriptor training flow chart

The specific steps in the training phase are as follows:

1. The n amplitude images are randomly selected from the image library.

2. Convert the dimensions of all images to 256*256, and convert their color space into lab space.

3. For each image, it is divided into 16*16 pixels of 256 small blocks. For each piece, find the pixel value of the most frequency. If a block with this pixel value is more than 5 blocks, a pixel value of this block is randomly selected as the

primary color of this block; otherwise, the pixel value with the most frequency will be the primary color of this block.

4. Use the k-means algorithm to train 256*n color values in step 3 to generate a palette containing k_c color values. That is, the color code (codebook). $C = \{c_1, ..., c_{kc}\}$.

The detailed process for generating the BOC descriptor is as follows:

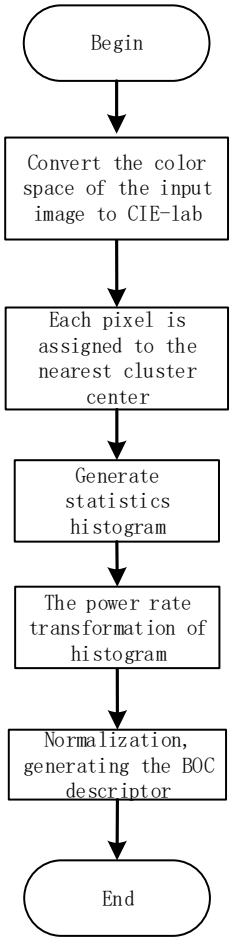

Figure 3. BOC descriptor generation flowchart

1. The initial BOC descriptor v is the zero vector of dimension k_c. k_c is the number of code words in the code.

2. Convert input image to CIE-lab color space.

3. Each pixel of the image is p in the color coding book to find the nearest code c_i. The corresponding statistical number $v_i + 1$; that is, is assigned to the nearest cluster. It is shown in Equation (6).

$$c(p) = c_i = \begin{cases} c_i \in codebook \mid \forall c_j \in codebook \setminus c_i, \| p - c_i \| \leq \| p - c_j \| \\ 1 \leq i \leq k_c, 1 \leq j \leq k_c \end{cases} \quad (6)$$

4. The statistical histogram is formed and the power rate transformation is adopted by Equation (7) to reduce the influence of noise.

$$V = [\sqrt{v_1}, ..., \sqrt{v_{kc}}] \tag{7}$$

5. Normalizing vector of L1 is adopted. It is shown in Equation (8).

$$x = \frac{x}{\sum_{i=1...k_c} x_i} \tag{8}$$

3. High dimension vector reduction

Kernel Principal Component Analysis (KPCA) is a nonlinear extension of the traditional principal component analysis by using kernel techniques. Compared with the principal component analysis, KPCA not only effectively captures the nonlinear characteristics and distribution of data, but also does not increase the computational complexity. Due to the large feature dimension of objects in practical problems, it is easy to reach thousands or even tens of thousands of dimensions. As a result, direct processing and storage of such data may result in system inefficiency. However, it is unnecessary to utilize all the features of the object for a specific application, so a large number of features cannot reflect the essence of the object. Sometimes, redundant features may become noise that affects recognition. KPCA can analyze the data, find out the most important elements and mechanisms in the data, and remove noise and redundancy, ultimately reducing the dimension of high-dimensional data.

In order to make the SD-VLAD and BOC descriptors suitable for large-scale retrieval, it is an essential step for SD-VLAD and BOC reduction.

The detailed steps to reduce the dimension of the local aggregations descriptor of d dimension for KPCA are as follows:

1. The n training vectors in the database are standardized so that all the vectors in each dimension are numeric and zero.
2. Calculate the kernel matrix according to the kernel function. The kernel matrix is represented by $n*n$ matrix K.
3. The eigenvalue decomposition of the nuclear matrix K after centralization is performed.
4. The eigenvectors are calculated and sorted according to the order of eigenvalues, and the former $d^{'}$ eigenvectors are retained.
5. The unit feature vectors and concatenate them into $d*d^{'}$ dimensional feature matrix.
6. The original local aggregation vector is multiplied by the characteristic matrix, and the eigenvector is obtained after the dimension reduction, and its dimension is $d^{'}$

With the reduction of dimension $d^{'}$ after dimensionality reduction, the dimensionality reduction error will increase continuously. The original BOF can also reduce dimension by dimension;0 however, the accuracy of vector query decreased sharply. When the dimension $d^{'}$ after SD-VLAD dimension reduction is 1/8 before dimensionality reduction, the loss of information is very little. The accuracy of the query result was basically the same as before the reduction.

4. Analysis of selected soccer video features

There are three main features of the soccer video retrieval system: the BOC description feature, the SD-VLAD description feature and the lens type feature.

The local feature used to generate SD-VLAD is the SURF feature (Speeded Up Robost Feature). It is shown in Table 1. SIFT performs best under the conditions of scale and rotation transformations. SURF has the best match performance under the brightness changes, shows superior fuzziness performance to SIFT, yet offers weaker changes in size and rotation than

SIFT and much weaker than SIFT in terms of unchanged rotation. It is worthwhile to buy a shortened retrieval time at the cost of a small amount of precision. Therefore, it is appropriate to select the SURF feature as the football video's local feature.

Table 1. Evaluation of local feature performance

Features	Time	Scale change	Rotate change	Fuzzy	Brightness change	Affine change
SIFT	Commonly	Best	Best	Commonly	Commonly	Preferably
PCA-SIFT	Preferably	Preferably	Commonly	Best	Preferably	Best
SURF	Best	Commonly	Preferably	preferably	Best	Preferably

The local feature is characterized by large number and big memory space, so it is unsuitable as a storage feature for mass image retrieval. By contrast, SD-VLAD may integrate with local features and compress the memory space under the premise of guaranteeing the precision. As a result, it is selected as football video's image feature. However, the image contents vary widely. SD-VLAD descriptors with the same parameters present significantly different recognition rate in different image regions. Thus, it is essential to select the most suitable parameters for football video images through experiments.

The SURF feature is the local descriptor vector generated by extracting the key points after transforming the image into a grayscale image. According to the SURF feature extraction process, SURF feature loses the color of the image, which cannot fully express the image information. Therefore, BOC feature description is introduced into the football video retrieval system to make the SD-VLAD feature and the BOC feature complement each other. This improves the query precision of the image. BOC feature's descriptor parameters also need to be set separately according to a specific region of the image, in order to play the role of BOC features.

Video lens type is one of the characteristics of football video, which reflects the content information of the image. The images of different lens types vary greatly in content. Therefore, adding the lens type as the image feature can further improve the query accuracy.

5. Experimental analysis and results

In order to detect the SD-VLAD and BOC descriptor in the influence of parameters on the descriptor identification, and select appropriate parameters, the paper uses the picture library on the search net to test it, and analyze the result of the experiment.

5.1. Index of retrieval evaluation

Image retrieval evaluation index is a set of objective numerical values, which is an accurate measure of the performance of image index and retrieval method. In the field of image retrieval, the most widely used evaluation index is precision and recall.

The precision is a retrieval process, and the relevant picture of the system return result is the proportion of all returned images. The recall is the proportion of related images in the whole image library. The accuracy of the retrieval is reflected by the precision, and the recall reflects the comprehensiveness of the retrieval. It is often necessary to set the weight of two ratios according to the requirements of the system.

However, it is difficult to describe the performance of complex system with the above two evaluation indexes. So, Jegou [13] proposes to use MAP (mean average precision) and average recall to evaluate the performance of the retrieval method.

The average precision rate is the area under the recall-precision curve. The larger the value obtains, the higher the retrieval quality is. This paper also employs the average precision rate to evaluate the performance of multi-feature fusion methods in queries.

The average recalling rate is the average of the ratio of the query results to the total number of returned results among the returned R results. This evaluation index should be determined by multiple queries. The average recalling rate curve can be obtained by setting different values of R. The farther the distance between the average recalling rate and the x axis, the better the retrieval algorithm performance.

5.2. Experimental environment

1. Hardware environment: Intel (R) Xeon (R) CPU 8 core, 16G memory

2. Operating system: Centos 5.6

3. Programming environment: Matlab 7.10.0, Codeblocks

5.3. Experimental data

The image database is used in soccer video key frame selection from the search net in the database. The experimental image library contains 8,862 photos of telephoto lenses, 8056 photos of non-telephoto lenses, 82 photos of the remote camera, and 51 photos of non-remote lens query. This query performance is evaluated using a MAP; the larger the MAP value, the better the query performance.

5.4. Experimental result

5.4.1. SD-VLAD descriptor and VLAD selection of clustering number

k is the number of clustering center set when the SD-VLAD and VLAD descriptors are generated. Figure 4 is the MAP graph of SD-VLAD and VLAD descriptors when the number of clustering center is k = {16,24,32,64}.

In Figure 4, compared with VLAD descriptors, MAP is 2% higher on average for SD-VLAD. According to the slope of the curve, it is evident that as the number of clustering center increases, the MAP values of the two descriptors increase continuously, yet with smaller improvement range. The larger the value of k, the larger the dimension of the generated descriptor, and the higher the cost of dimension reduction. Although the dimension of descriptors reduces as the value of k becomes smaller, the retrieval performance of the aggregated descriptor will also be affected. Due to the different characteristics of the image, the image retrieval system focuses on different points. It is necessary to strike a balance between the query speed and query precision for different image libraries, so as to select the optimal k value through constant experiments and adjustment. For football video image library, the best k value is generally 24.

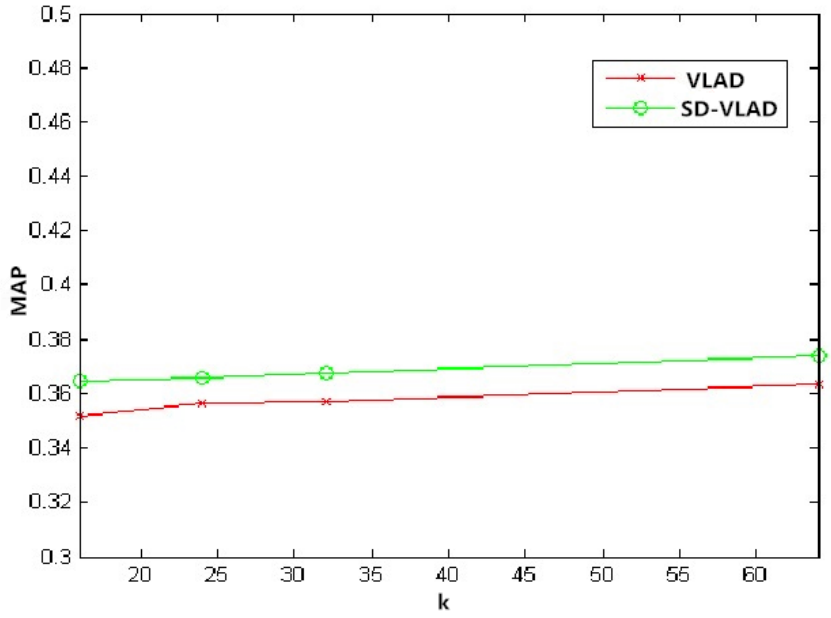

Figure 4. SD-VALD and VLAD query precision comparison diagram

5.4.2. The BOC descriptor selects the number of clusters

The number of color clustering center k is a key parameter for generating a BOC descriptor. It directly affects the BOC

descriptors' query performance and query time. From the data of Table 2, the larger the value of k, the better the average query precision of BOC descriptor and the longer the feature extraction time becomes. Although BOC descriptors' average precision rate reduces as the value of k becomes smaller, the feature extraction time is shortened correspondingly. Same with SD-VLAD; k value should be set experimentally by striking a balance between speed and the average query precision rate for different image libraries. For football video photo library, we generally choose k =256 as the best value.

Table 2. The mean average precision and feature extraction time of BOC descriptor.

k	32	64	128	256	512
MAP	0.279	0.337	0.364	0.382	0.391
Extraction time /ms	12	15	21	31	55

6. Conclusions

Based on the analysis of video characteristics of football, this paper puts forward a description of SD-VLAD feature descriptor, BOC feature descriptor and lens type. The effect of cluster center parameters on the query performance of the corresponding descriptor is tested by experiment. The picture content of different fields is different. Fixed parameter settings cannot guarantee the query accuracy of SD-VLAD and BOC descriptors in various fields. Therefore, it is necessary to select the best parameters for soccer video images by experiment.

References

1. E. Andreas, B. Herbert, "SURF: Speeded Up Robust Features", *Computer Vision and Image Understanding*, vol.110, no.3, pp. 346-359, 2008
2. O. Aude, 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
3. O. Aude, "Building The Gist of A Scene: The Role of Global Image Features in Recognition", *In: Proceeding of Progress in Brain Research*, pp.23-36, 2006
4. G. David, "Distinctive Image Features from Scale-invariant Key Points", *International Journal of Computer Vision*, vol.60, no.2, pp. 91-110, 2014
5. M. Douze, "Evaluation of GIST Descriptors for Web-scale Image Search", *In: Proceeding of the ACM International Conference on Image and Video Retrieval*, pp. 1-8, 2009
6. P. Florent, M. Thomas, "Improving the Fisher Kernel for Large-scale Image Classification". *In: Kostas Daniilidis, Petros Maragos, Nikos Paragios. Proceeding of the 11th European Conference on Computer Vision (ECCV)*, pp. 143-156, 2010
7. H. Jegou, D. Matthijs, "Aggregating Local Descriptors into A Compact Image Representation", In*: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp.3304-3311,2010
8. P. James, C. Ondrej, 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
9. S. Josef, 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. M. Krystian, "A Performance Evaluation of Local Descriptors", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.27, no.10, pp.1615-1630, 2005
11. J. Michael, "Color indexing", *International Journal of Computer Vision*, vol.7, no.1, pp.11-32, 1991
12. Y. Pu, "A Review of Research on The Key Technology of Content Based Video Retrieval", *Information Science*, vol.28, no.3, pp.464-469, 2010
13. K. Timos, "The R+-Tree: A Dynamic Index for Multi-Dimensional Objects", *In: T. M. Vijayaraman. Proceeding of the 13th International Conference on Very Large Data Bases*, pp. 507-518, 1987
14. K. Yan, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors", *In: Proceeding of the IEEE Transactions on Computer Vision and Pattern Recognition*, pp. 506-513, 2004

Yang Han received his M.A degree from University of Canberra. He is a lecturer in Heilongjiang University. His research interests include physical education, sports training and sports humanities.