International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.1 (2016), pp.85-94
http://dx.doi.org/10.14257/ijsip.2016.9.1.08

Camera Lens Detection Algorithm based on the Dominant Color Image in Soccer Video

Xiang Zhai

Department of Physical Education, Northeast Forestry University,
Harbin 150040, China
597390959@qq.com

Abstract

Key events in football videos are fetched on the basis of accurate detection of shots. A quick and effective algorithm for detecting football video shots has very important meanings. Here it focuses on features of football videos by combining with a few basic shot detection methods to propose one shot detection method applicable for football videos. The experiment shows the feasibility of the proposed algorithm, and the experimental data shows that the algorithm has a good effect for soccer video shot detection and key frame extraction

Keywords*: Video retrieval, Shot segmentation, Soccer video, MPEG, Dominant Color*

1. Introduction

Compared with other video types, football videos have own characteristics. In the following we summarize difficulties in football video shot boundary detection and discuss how to deal with such problems in determining boundaries [1-2].

In football videos, strong color correlation exists among different shots, which is not found in general videos. That's because in lots of consecutive shots, there's only one main color background, like green football pitch. So the shift of lens may not lead to obvious frame differences. That presents challenges to the traditional algorithms which detect shot boundaries by according to big frame differences [3].

In this case, we raise multiple thresholds to adapt to the detection of shots of different types. We use major color ratio to detect the visibility of the field. When the principal color ratio is high, the pitch is visible and we'll use smaller threshold; otherwise, we'll use bigger threshold.

One set of complete football video includes suddenly shifted shots and also gradually changed shots. Traditional shot detection methods didn't do well in detecting gradually changed shots. Hence, we apply variable step algorithm to implement the boundary detection of different shots [4-5].

Football videos used here are stored in MPEG format. If they're all decoded, it requires tremendous time and storage space. We can actually decode part of them to get the DC image sequences. Then with the step size detection algorithm for DC image sequences, we can realize the detection of shot boundaries.

2. Variable Step Detection Algorithm based on DC (Dominant Color) Images

2.1. Selection of Features

The shot detection requires firstly selecting a suitable feature value as for feature extraction. In light of difficulties mentioned above, we choose the two properties:

ISSN: 2005-4254 IJSIP

2.1.1Color Histogram Frame Difference D (i, k): We choose color histogram as one eigenvalue for the reason that it is not quite sensitive to small motions of objects and cameras. The color space is selected YUV space. The similarity between two frames is measured in the expression of luminance Y histogram frame differentiation method. By equation (1), we can calculate the similarity between the ith and the kth frame; then normalize the result to [0-1].

$$D(i,k) = \frac{1}{(M \times N)} \sum_{j=1}^{n} H_i(j) - H_k(j)$$

(1)

Where: n is the histogram of the interval number, M is the height of the image, N is the width of the image.

2.1.2 Main frame color ratio G_i : This paper introduces the main color pixel ratio characteristics of G_i image, the color image is the main area for the frame in the proportion. In the soccer video G_i is the ratio of grassland area accounted for the entire image.

2.2. Fetch of DC Image Sequences

At first we decode football videos compressed in standard MPEG format as to fetch DC image sequences, which are used as original data for variable step size shot detection. DC images have two obvious merits:

They contain all basic global information of original images;

Compared with full images, DC images scale down apparently, good for more rapid compression processing.

The method of DC extraction from image sequences [6-9]:

For one 8*8 image block, the relationship between a 2-dimension discrete cosine transform (DCT)'s DC coefficient c (0,0) and its original 8*8 pixel is shown as follows:

$$c(0,0) = 1/8 \sum_{X=0}^{7} \sum_{Y=0}^{7} f(x,y)$$

(2)

For one M*N image, replace its pixel value with DC coefficient in each block; the image is compressed to one DC image at the compression ratio 64:1. It shown in Figure 1.

(a) The original frame image (b) extraction of DC image

Figure 1. DC Image Extraction

For both frame I, it can be directly from the DCT coefficients extraction DC coefficient C (0, 0), then divide by 8 to get.

For both frame P and B, since what they transfer is DCT coefficient of residual error D after prediction or interpolation, the actual DCT coefficient can be obtained by motion compensation. We take frame P for instance. It shown in Figure 2. Set the current block P_{cur}, remainder error D_{cur} and motion vector $(\Delta x \Delta y)$. P_{ref} is one 8*8 block used for prediction in reference frames, which we call reference block. Then, we have:

$$DC(P_{cur}) = DC(P_{ref}) + DC(D_{cur})$$
(3)

Hence, as long as we get P_{ref}'s DC component, we can get $DC(P_{cur})$ eventually. P_{ref}'s position can be obtained according to P_{cur}'s coordinate vector and motion vector $(\Delta x \Delta y)$, which is covered with four adjacent P_1, P_2, P_3, P_4 blocks; h_i and w_i refers to respectively height and width of intersecting rectangle by P_{ref} and P_i. Apparently,

$$h_1 = h_2 = 8 - h_3 = 8 - h_4$$
$$w_1 = w_3 = 8 - w_2 = 8 - w_4$$
(4)

Since DCT transform is linear transform, $DC(P_{ref})$ can be acquired with P_1, P_2, P_3, P_4's DCT coefficient; but it requires huge computational work. In this regard, we can apply the approximation algorithm proposed by Yeo [10], utilizing P_{ref}'s pixel percentage of P_i as weighted value to totalize P_i's DC coefficients and make the sum as $DC(P_{ref})$'s approximate value: It shown in the formula (5):

$$DC_{ref} = \frac{1}{64}\sum_{i=0}^{3} h_i w_i DC_i$$
(5)

using bidirectional prediction block for B frame in the DC coefficient, it can according to the similar method with two predicted macro block DC coefficient calculated average approximation.

$$DC(B_{cur}) = [DC(B_{pre}) + DC(B_{back})]/2$$
(6)

Where, $DC(B_{pre})$ is the forward prediction reference component of the DC block, $DC(B_{back})$ is the DC component to forecast reference block.

With the above method, we can decode MPEG video sequences to DC image sequences and regard them as original images for shot detection. In the following section, all frame images used here are DC images. For the convenience of expression, they are named collectively frame images.

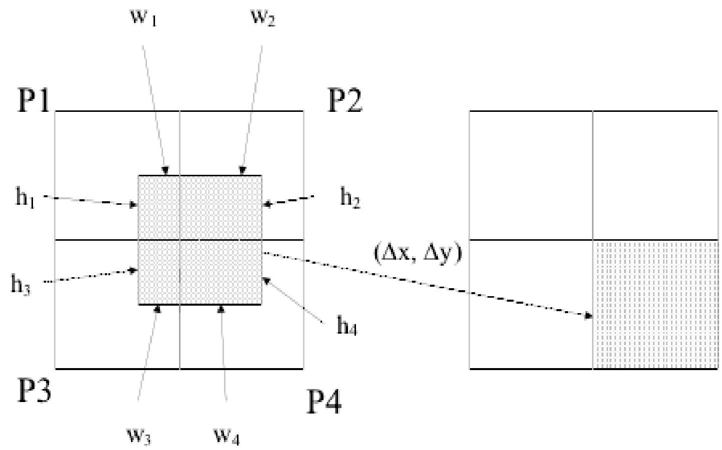


Figure 2. Reference Block and Motion Vector

2.3. Selection of Threshold Value

As per the different percentage of the main color area in frame images, we classify shots [11-12]. In football videos, the main color area in frame images is green land. Based on the proportion of green area to the whole frame image, we can define three shots:

(1) Long shot;

(2) mmedium shot;

(3) Non-field shot or close shot. See Figure 3 for details.

(a) the long shot, (b) in the distance of lens

(c) the close-up (d) field region lens

Figure 3. The Definition of Different Shot in Soccer Video

In the picture, long shots cover the most extensive grass lawn pixels; in medium shots, there's partial green area; while in close shots or non-field shots, no grass area is seen. So it's very effective to make initial judgment of shot types by referring to the proportion of grassland area [13-14].

After analyzing pixels of frame images, we find that green color component in football playground is bigger than red and blue component. We'll get accurately the field area provided that the difference is calculated between green component and the other two.

$$f(x, y) = \begin{cases} G, if \quad g(x, y) > r(x, y) \; and \; g(x, y) > b(x, y) \\ N, other \end{cases}$$

(7)

Where: r (x, y), g (x, y), b (x, y) denote the pixels (x, y) of each color component values. G said the pixel point marked as green, N marked non-green.

Each image pixel marked as the number of green is denoted by m, the total number of pixels per frame is M. Then the field region in each frame ratio:

$$G_i = \frac{m}{M}$$

(8)

It can be judged that whether the field is visible in frame images. After analyzing plenty of football videos, we selected $T_1 = 0.15$. Looking back to different features of football video shots, we chose different threshold values to detect shots. When the field is invisible in frame images, the frame difference is huge; then we select bigger threshold to make shot detection of T_H^{high}, T_H^{low}; otherwise, we choose smaller value to do the same work of T_L^{high}, T_L^{low}.

3. Experimental Analysis and Results

This experiment is in the Windows XP operating system, Visual C++ 6 development environment to complete. The experiment used 5 field compression for soccer video clips of MPEG-1 as the test data, respectively denoted as test1, test2, test3, test4, test5.

3.1. DC Image Sequences

3.1.1. Decoding of DC image sequences. First of all, football videos in standard MPEG compression format are partially decoded to resave as DC image sequences. Table 1 lists out the storage space size of five experimental data after partial and full decoding. Clearly, DC image sequences take up the least storage space, only 343M; while full image sequences consume as much as 44.8G and its decoding speed is rather slow, not helpful to the shot detection afterwards.

Table 1. Compared with DC image, MPEG Video and Storage Space of the Whole Image

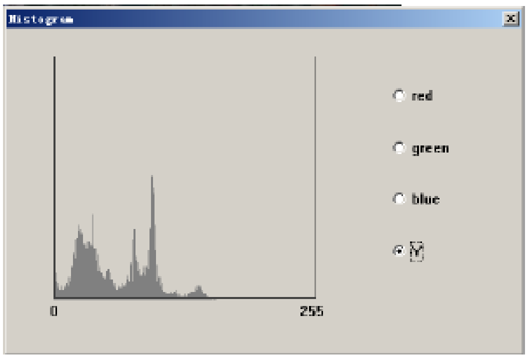
	MPEG-1	DC image sequence	The whole image sequence
Test1	118.6M	70.8M	9166.4M
Test2	125.M	73.3M	9591.1M
Test3	109.6M	61.3M	8091.3M

Test4	134.6M	79.6M	10409.6M
Test5	97.4M	58.0M	7534.2M
Total	579.3M	343M	44792.7M

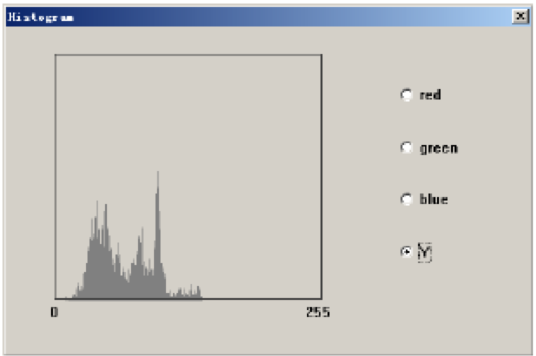
3.1.2. DC Image and the Image Comparison: After decoding football videos in standard MPEG compression format, we can fetch DC image sequences and use them as original data for variable step size shot detection. Figure 4 compares full images and DC images. From it, we see DC images contain all basic global information of the original images. Luminance histograms of two images are generally consistent. Besides, DC images shrink apparently, characteristic of quick compression processing. Figure 5 gives the comparative histograms of full images and DC images in 1000 frames of one football video fraction. In it, the histogram of DC images changes sharply, because DC images are extracted with the approximation algorithm mentioned by Yeo [15]. However, boundaries of shots are clearly represented.

(a)The original frame image

(b) Extraction of DC image

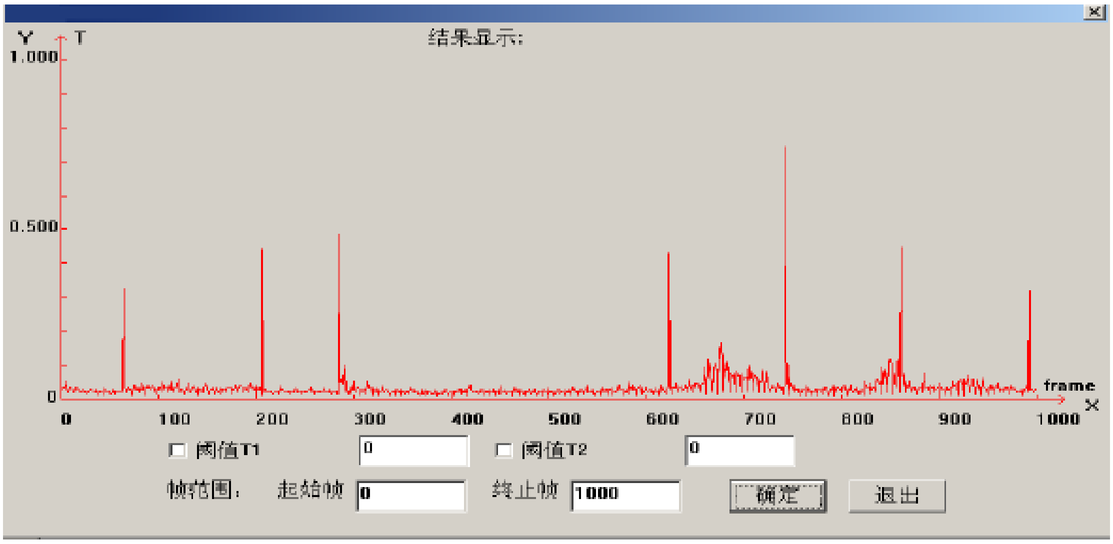


(c) Image luminance histogram

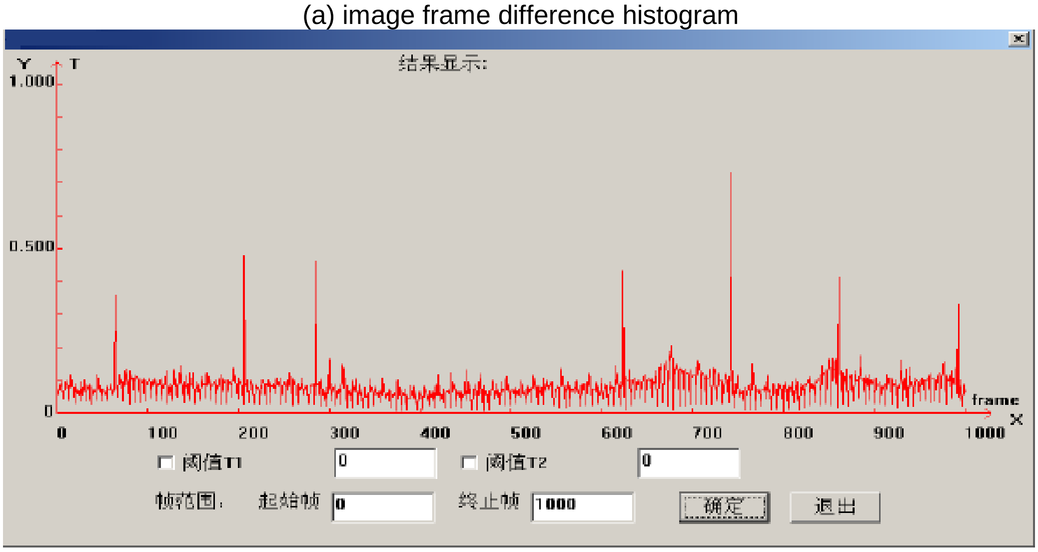


(d) DC image brightness histogram

Figure 4. Comparison of Frame Image and DC Image



(a) image frame difference histogram



(b) DC image frame difference histogram

Figure 5. Comparison of The Whole Image and DC Image Frame Difference Histogram

3.2. Selection of Threshold Value

Fig. 6 is frame brightness histogram of one football video clip. We note that in position ① and ②, lens had sudden shifts. When such shifts occurred, big differences appeared in the column diagram of frame difference brightness. After analysis of original images, we noticed that the field is not observed in shots in ① and there're big frame differentials; so we need to set bigger threshold for the detection; in shots in ②, the field is visible; since the field color is principal, no big differences occurred to the histogram of frame images; if we continue using threshold in ①, there will cause lots of leak detection. To cope with ②, we choose smaller threshold value to implement shot detection.

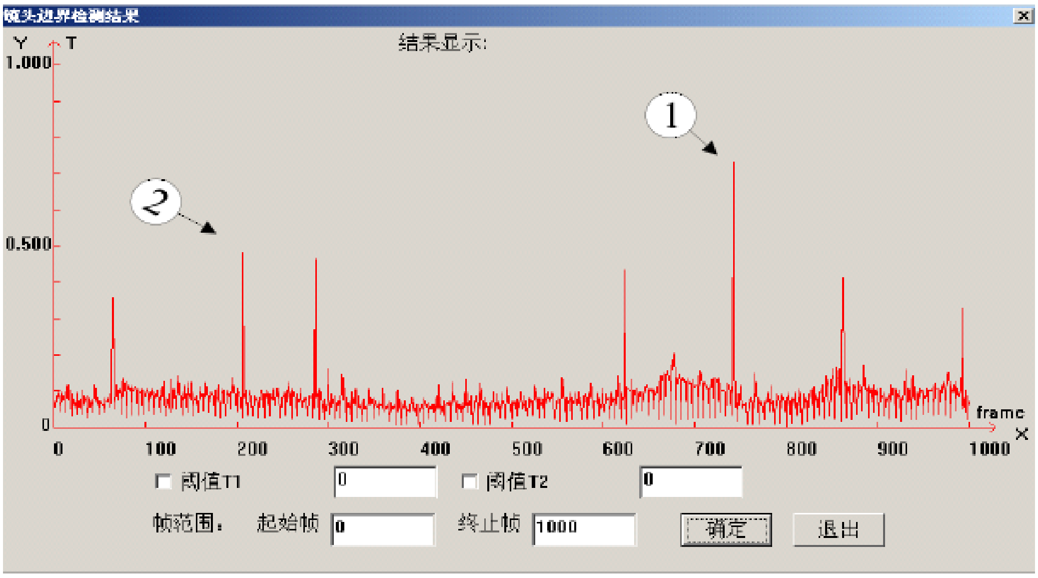


Figure 6. Frame Brightness Histogram

Through the analysis of a large number in soccer video, threshold in this paper: T_H^{high}, T_H^{low} =(0.65,0.15). T_L^{high}, T_L^{low} =(0.3,0.12)

3.3. Detection Results of Shots in Football Videos

The performance of shot boundary detection algorithm is often used to Recall and Precision to represent:

$$Recall = \frac{N_c}{N_c + N_m} \times 100\% \quad (9)$$

$$Precision = \frac{N_c}{N_c + N_f} \times 100\% \quad (10)$$

Where, N_c is the number of correctly detected lens, N_m is lens of the missing number, N_f is lens number of false detection

Set well the threshold and make step size mystep=30; we can perform variable step size shot detection based on DC images for the selected football videos. The results shown in Table 2.

For five groups of experimental data, we utilized the full image shot detection based on histogram by Peng. The results shown in Table 3.

Table 2. The Results of the Lens Detection of Variable Step based on DC Image

	N_c		N_m		N_f		Recall (%)	Precision (%)
	Mutation	Gradient	Mutation	Gradient	Mutation	Gradient		

Test1	113	14	3	2	7	2	96.1	92.9
Test2	109	8	4	1	4	2	95.7	94.9
Test3	120	10	2	1	5	1	97.7	95.4
Test4	126	12	3	2	5	1	96.4	95.6
Test5	95	8	2	1	3	2	97.1	95.1
Total	563	52	14	7	24	8	96.6	94.8

Table 3. Lens Detection Results based on the Histogram Graph

	N_c		N_m		N_f		Recall (%)	Precision (%)
	Mutation	Gradient	Mutation	Gradient	Mutation	Gradient		
Test1	113	14	14	5	8	2	85.0	92.1
Test2	109	8	12	4	5	3	96.3	93.2
Test3	120	10	15	4	4	3	85.4	94.6
Test4	126	12	17	5	5	2	84.1	94.9
Test5	95	8	9	3	3	3	88.3	94.2
Total	563	52	67	21	25	13	85.7	93.8

As shown in Table 2 and 3, Peng's histogram-based shot detection algorithm got higher precision rate, even though recall ratio is not satisfactory. That is because the main color in football videos is green color of the pitch, resulting in slight histogram difference between frames. But if by the traditional shot detection method with bigger threshold value, a great deal of shots will be missing. Likewise, if with smaller threshold value, there must be wrong detection of shots. The proposed variable step size shot detection algorithm based on DC images can achieve both higher recall ratio and precision rate, sufficing for the requirement of detecting football video shots. On the other hand, Peng's shot detection solution needs to decode videos to full images, which will consume abundant storage space and time, impossible to improve the shot detection efficiency. By contrast, the new strategy based on DC images requires partial decoding, which saves plentiful storage memory and time.

Experimental findings reveal that the football video shot detection method based on DC images reaches better recall and precision rate, however, it has shortcomings:

The selection of threshold value is of certain subjectivity as the choice is made after analysis of numerous football video shots;

The precision of detecting gradually changed shots requires improvement. In light of no big inter-frame differences, there are some difficulties in detecting gradual varied shot detection, especially when gradually changed shots appear between two long shots or midrange shots.

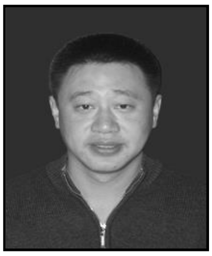
4. Conclusion

In the paper, firstly we decoded football videos in standard MPEG compression format to fetch DC image sequences; based on the main color ratio, we detected the visibility of field in frame images. Secondly, with variable step algorithm, we realized the detection of gradually and suddenly shifted shots in football videos. Thirdly, one key frame extraction algorithm was proposed based on principal color ratio. Through experiments, the new algorithm proved its feasibility. The results showed that the method made better effects in detecting football video shots and fetching key frames.

References

[1] Z. Bin, "Research on analytical method for soccer video content annotation", Jilin University, **(2008)**.
[2] Q. Zechao, "Research and implementation of soccer video retrieval. Nanjing University of Science and Technology", **(2008)**.
[3] C. Yongping, "Research on Key Technologies of typical events detection in soccer video", Chongqing University, **(2009)**.
[4] Q. jade, "Classification of shot detection and motion activity based video", Beijing University of Posts and Telecommunications, **(2013)**.
[5] Z. Yuan, "Research and analysis of tennis video semantic based on multi-modal", Nanjing University of Science and Technology, **(2013)**.
[6] O. Bao and L. Guan, "Scene change detection using DC coefficients", Image Processing, **(2002)**, pp. 421-424.
[7] O. Bao, "Scene change detection using DC coefficients", IEEE ICIP, **(2002)**, pp. 421-424.
[8] B. Yeo and B. Liu, "Rapid scene analysis on compressed video", IEEE Trans. Circuits Systems Video Technol., vol. 5, no. 6, **(1995)** , pp. 533-544.
[9] K. Shen and E. J. Delp, "A Fast Algorithm for Video Parsing Using MPEG Compressed Sequences", Proceedings of International Conference on Image Processing (ICIP'96), Lausanne, **(1996)**.
[10] B.-L. Yeo and B. Liu, "On the Extraction of DC Sequence from MPEG Compressed Video", ICIP, **(1995)**, pp. 2260-2263.
[11] L. Haibin and Z. yi, "An efficient video shot boundary detection algorithm", China Journal of image and graphics, vol. 4, no. 10, **(1999)**, pp. 805-810.
A. Ekin, AM Tekalp and R. Mehrotra, "Automatic soccer video analysis and summarization", IEEE Trans. Image Processing, vol. 12, no. 7, **(2003)**, pp. 796-807.
[12] C. W. Ngo, T. C. Pong and Z. Hongjiang, "On clustering and retrieval of video shots through temporal slices analysis[J]", IEEE Transactions on Multimedia, vol. 4, no. 4, **(2002)**, pp. 446-458.
[13] S. Takagi, S. Hattori and K. Yokoyama, "Sports video categorizing method using camera motion parameters", Proceedings of the Conference on Multim edia and Expo 2003, Baltimore, Maryland, IEEE Computer Society, **(2003)**, pp. 461-464.
[14] X. Peng, X. Lexing and S. F. Chang, "Algorithms and system for segmen -tation and structure analysis in soccer video", IEEE International Conference on Multimedia and Expo 2001, Tokyo, Japan, IEEE Computer Society, **(2001)**, pp. 721-724.

Author



Xiang Zhai, He received his B.S degree in physical education from Sports Science Institute of Harbin Normal University. He got his M.S degree in physical education from Shenyang Sport University. He is a lecturer in Department of Physical Education, Northeast Forestry University. His research interests include sports sociology.

,