International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.2 (2016), pp.25-34
http://dx.doi.org/10.14257/ijsip.2016.9.2.03

Sports Games Video Segmentation based on Adaptive

Wei Liu

Physical Educational Department of Heilongjiang University,
Harbin 150001, china
liuwei@hlju.edu.cn

Abstract

Based on analyzing the related technologies of the original video retrieval, we proposed an algorithm of adaptive dual-threshold shot boundary detection for the game video of TV. At the beginning, we convert each frame to the HSV space, set high and low ratio of threshold according to the frame difference. Then, the average value of frame difference in the sliding window is calculated to receive the dual-threshold. Finally, the shot and the gradual are detected by comparing the frame difference and threshold.

Keywords: *sport video analysis, video retrieval, shot detection, adaptive dual-threshold*

1. Introduction

Sport video retrieval based on contents is done from tremendous video data according to video contents and contexts [1-3]. The key to do that is split shots which are joined up by various compiling tools, *i.e.,* video segmentation, also called shot edge detection. The paper introduces some common detection methods for shot boundaries and compares them. On that basis, it proposes the adaptive dual threshold detection method in accordance to the viewing transformation feature of volleyball videos [4-5]. To enable gentle viewing transformation, it gives out good threshold coefficient for better effects and detection efficiency. Besides, it mentions some key techniques used in the implementation such as HSV color space model and the transformation between color space models. In the end, it shows experimental results of the proposed approach and proves its effectiveness and practicability [6].

2. Common Shot Detection Methods

For telecast sport match videos, they are put together by some video clips. In terms of shot cuts spliced in different manners, we can have abrupt transition and gradual transition shot [7]. The former refers to direct connection of two shots, without any photographic editorial effect nor any time delay; the latter involves shifting gradually from one to another shot, without noticeable visual skipping. So far the general video gradual transition includes wipe change and dissolve. What we talk mostly in sport videos is wipe transition, that is, use one LOGO icon to scan from one side to the other side of image. Wipe change is mainly used to play back wonderful shots. It is shown in Figure1-2.

ISSN: 2005-4254 IJSIP

Figure 1. Abrupt Transition

Figure 2. Gradual Transition

2.1. Abrupt Transition Shot Detection

Abrupt can be direct switching or transition. Direct switching can enable direct jumps of image's scenario and movements, no temporal or spatial transition found in the intersected part of two shots. In general, frames within the same shot are slightly different from one another, but they differ a lot in different shots, which is easy for shot detection. At present, the common abrupt shot change detection methods include template matching algorithm, histogram method, edge-based method and model-based method.

2.1.1. Template Matching Algorithm [8]: It is renamed pixel-based matching method, which regards the summation of absolute values of two frames' pixel differences as inter-frame difference. Frame difference calculation is shown in a formula 1.

$$d(I_i, I_j) = \sum_{x=0,\, y=0}^{x<M,\, y<N} |I_i(x,y) - I_j(x,y)| \tag{1}$$

2.1.2. Histogram Method [8-9]: The method uses statistical values of pixel illuminance and color, no relation with pixel's position information. It's more resistant to noises than template matching method. Hence, it is used most widely to calculate histogram differences between frames. The principle of realization is: suppose the color space is quantified to n handles; H is value of normalized histogram of the i frame in the k handle. Frame difference calculation i shown in a formula 2.

$$d(I_i, I_j) = \sum_{k=1}^{n} | H_{ik} - H_{jk} | \tag{2}$$

2.1.3. Edge-based Method: After video analysis, we find that when lens change, new frame boundaries should keep away from the location of edges of those missing frames; likewise, edges of missing frames should be away from the location of new emerging frames. Based on that characteristic, we can detect shot boundaries.

2.2. Gradual Transition Shot Detection

When gradual transition exists, the frame image change transists slowly. Inter-frame difference grows bigger, but there is no obvious peak, instead, a "plateau area" lasting for some time, for that, the abrupt detection method is not applicable. To address it, Zhang *et* *al.*, [10]. suggested dual threshold comparison technique. Based on video features, they set high threshold Th and low threshold Tl to compute the inter-frame $d(I_i, I_j)$; then, they compared with predefined threshold value. If Th< $d(I_i, I_j)$ <Tl, it's start frame with potential gradual transition and necessary to calculate the cumulative differentials of its frame difference for every subsequent potential transformation Ac(i)+= $d(I_i, I_j)$; if it suffices Ac(i)>Th, it's believed gradual transition exists; if $d(I_i, I_j)$ <Tl, it's believed the gradual transition is over. The method is much effective for the detection of gradual transition. But lens' slow movements have the above feature too, false detection is easily caused.

Fernando *et. al.*, [11]. built mathematical model to address changing features of grey scale and color space after analyzing the dissolve of gradual transition shots. The principle is: suppose two images A and B, which decide jointly but not evenly video signals. We can understand the dissolving procedure of gradual transition in this way: image A deduces from 100% to 0 and image B increases from 0 to 100%. For fade-in, it doesn't consider image A; while for fade-out, it doesn't consider image B. We cite an example to show the fade-in process, as put in equation 3:

$$S_n(x,y)=\begin{cases} f_n(x,y) & 0\le n\le L_1 \\ \left[1-\left(\dfrac{n-L_1}{F}\right)*C+\left[\dfrac{n-L_1}{F}\right]*g_n(x,y)\right] & L_1\le n\le (L_1+F) \\ g_n(x,y) & (L_1+F)\le n\le L_2 \end{cases} \tag{3}$$

3. Common Color Models

Color model is known as color space or color system. Common color space is RGB, HSV *etc*. It's used to illustrate colors in a generally acceptable manner under some standards.

3.1. RGB Color Model [12]

In RGB model, each color appears in the original spectral component of red, green and blue. The model is built on Cartesian coordinate system. The color subspace discussed here is a cube as seen in Figure 1.

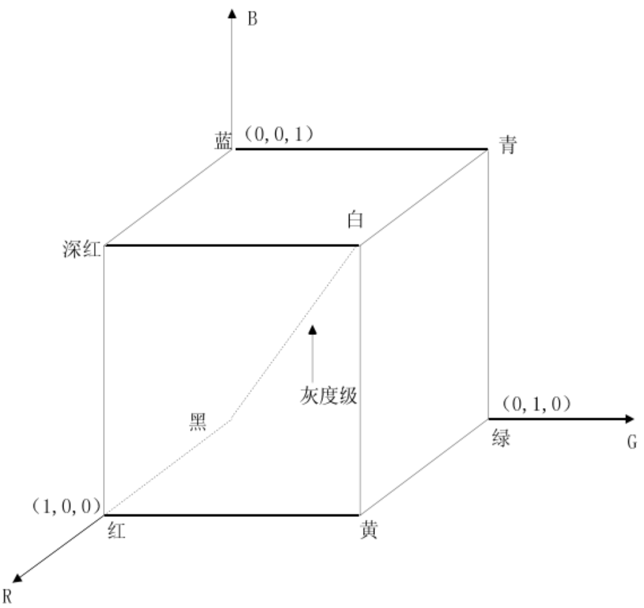

Figure 1. RGB Color Model

3.2. HSV Color Model [13]

This model is based on perception. It can be represented by three attributes: hue, saturation and brightness. Regarding human's perception and identification of colors, the model is rather suitable. HSV color model is geometrically one cone subset in the cylindrical coordinate system.

In the center of the top surface of the cone, s=0, v=1, h is not defined, on behalf of the white. At the apex of the cone, v=0, h and S is not defined, on behalf of black. Where H is the basic attribute of color, the range is 0 to 360 degree, the size is determined by the rotation angle of V axis; S refers to the color purity, the purity of higher saturation of color is higher, which ranges from 0 to 1; V is the brightness of the color, the range is from 0 to 1.It is shown in Figure 2.

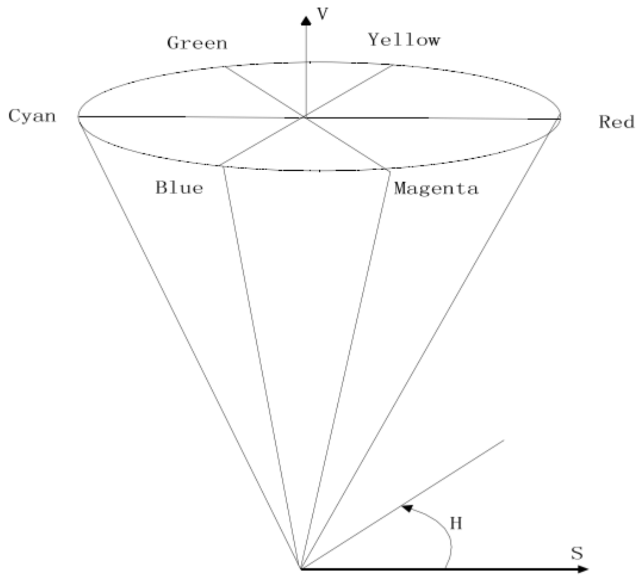

Figure 2. HSV Color Model

3.3. RGB-HSV Color Model Transformation

HSV color model is most similar to human's visual perception. The similarity distance between any two colors can be expressed with their positions in HSV space. It is also a consecutive color space. So it's quite suitable for the comparison of image similarities based on color. The design algorithm here adopts HSV color model. The transformation from RGB space to HSV space is depicted in equation 4-6.

$$H = \begin{cases} \arccos \dfrac{(R-B)+(R-G)}{2\sqrt{(R-G)^2+(R-B)(G-B)}} & B \le G \\ 2\pi - \arccos \dfrac{(R-B)+(R-G)}{2\sqrt{(R-G)^2+(R-B)(G-B)}} & B > G \end{cases} \tag{4}$$

$$S = \frac{\max(R,G,B)-\min(R,G,B)}{\max(R,G,B)} \tag{5}$$

$$V = \frac{\max(R,G,B)}{255} \tag{6}$$

4. Performance Evaluation of Shot Detection

So far, there has been no uniform solution to solve the problem regarding performance evaluation of shot detection.

Firstly, the appraisal of the goodness or badness of one algorithm is done as per performance indicators acquired after some standardized data. In image field, there're lots of standard image materials like Lena; while in video field, there has no generally accepted standard. So for each algorithm, its experimental results are inevitably one-sided. To make one algorithm universal and indicate problems qualitatively, videos used in the experiment must include a certain number of various change shots and that the algorithm can be convincing.

Secondly, we can base on one formalized equation to calculate the goodness or badness of one algorithm. Content-based video retrieval technique is hot topic raised recently. So far there has been no unified evaluation criteria to measure the shot detection effect. There're two existing standards which are publicly used and accepted. Of them, one uses recall ratio and precision rate as indicators in the multi-media information retrieval. It is shown in equation 7.

$$recall = \frac{N_{correct}}{N_{correct} + N_{miss}} \quad precision = \frac{N_{correct}}{N_{correct} + N_{false}} \tag{7}$$

Where, $N_{correct}$ is the quantity of shots which are correctly detected; N_{false} is the quantity of shots which are falsely detected; N_{miss} is that of shots missing for detection. The other is omission ratio and fall-out ratio, implying the percentage of missing detected shots and falsely detected shots against the total shots. It is shown in equation 8.

$$\eta_{miss} = \frac{N_{miss}}{N_{correct} + N_{miss}} \quad \eta_{false} = \frac{N_{false}}{N_{correct} + N_{false}} \tag{8}$$

Where , η_{miss} is missing rate, η_{false} is error rate.

The above two standards are mutually restrictive and contradictory. For different shot detection methods, although the detection of change shots cannot reach the best result, it's possible to choose the fittest solution as per different purposes and usages, to strike a balance between those indicators.

5. Shot Boundary Detection based on Adaptive Dual Threshold

5.1. Detection of Dual Threshold Shot Boundaries

Firstly, we quantify in unequal intervals the three color components H, S and V as per human's sensory ability to colors and visual discernibility and divide hue H, saturation S and luminance V to respectively 8, 3 and 3 portions. RGB space is transformed to HSV by the equation4-6. We use expression 9-12 to quantify different color ranges.

$$h' = 180 \times h / \pi \tag{9}$$

$$H = \begin{cases} 0, if\ h' \in [316, 20) \\ 1, if\ h' \in [20, 40) \\ 2, if\ h' \in [40, 75) \\ 3, if\ h' \in [75, 155) \\ 4, if\ h' \in [155, 190) \\ 5, if\ h' \in [190, 270) \\ 6, if\ h' \in [270, 295) \\ 7, if\ h' \in [295, 316) \end{cases} \tag{10}$$

$$S = \begin{cases} 0, if\ s \in [0, 0.2] \\ 1, if\ s \in [0.2, 0.7] \\ 2, if\ s \in [0.7, 1] \end{cases} \tag{11}$$

$$V = \begin{cases} 0, if\ v \in [0, 0.2] \\ 1, if\ v \in [0.2, 0.7] \\ 2, if\ v \in [0.7, 1] \end{cases} \tag{12}$$

Where, h, s, v were from RGB space to HSV space value. Have been quantified after H, S, V three components, you can put three weight value vector synthesis into one-dimensional. It is shown in equation13.

$$L = HQSQV + SQV + V \tag{13}$$

Based on the above equation13, we know that L is between [0, 71]; subspace label L to which any pixel's color value (h, s, v) belongs can be obtained. Assume one frame image has UxV pixels and color value of each pixel is L. Then we can get its color histogram $h_i(1)$. It is shown in equation 14.

$$h(i) = \sum_{p=1}^{U \times V} L \quad 1 \le i \le 200 \tag{14}$$

Secondly, with the acquired frame image histogram, we can calculate inter-frame difference. Here we use directly the simplest absolute value of two histograms as inter-frame difference. We give out two definitions: adjacent inter-frame difference and spaced inter-frame difference. The former refers to differentials between two neighboring frame images in the video. It is shown in equation 15.

$$D_{i,i+1} = \sum |(h_{i+1}(l) - h_i(l))| \tag{15}$$

Finally, we use the adjacent inter-frame difference Di got from equation 15 to compare separately with Th and Tl. If $D_{i,i+1}$ >Th, it's judged big visual changes between two adjacent frames, which is considered as Abrupt transition shot edge; if $D_{i,i+1}$ <Tl, it's believed no visual variations between them, which is considered as shot inner; if Tl< $D_{i,i+1}$ <Th, with the current frame i as potential abrupt start frame, we use the spaced

inter-frame difference acquired by equation 15 to compare with high threshold Th; if $D_{i,i+1}$ >Th, and two neighboring inter-frame difference meets $D_{i,i+1}$ <Tl, the following nth frame is thought as the end frame of this abrupt change.

5.2. Acquisition of Adaptive Threshold Value

The simplest and commonest detection method of shots is to preset two fixed thresholds as per experience, *i.e.,* global fixed threshold; then, by comparing adjacent frame differences between abrupt and gradual transition shot edges with two fixed thresholds, it judges whether it's shot edge. However, at different moments of videos, the illumination, noise and motion are not identical, even deviations of inter-frame within shots. As a result, there is leak detection or wrong detection by the algorithm. Meanwhile, different shot switching shows different features. It's not good to choose threshold as per previous experience. So we decide to use the method based on adaptive dual threshold [14-15].

In the beginning, we set one sliding window w, whose size is decided by the number of frames within one shot. Since there're many shot switches and slow-motion playback in volleyball match videos, the value of w can't be too big.

Since within one shot, the illumination and noises won't vary a lot, so the difference value between adjacent frames in one shot will be much close. Based on that trait, we can have inter-frame difference mean in one window. With mean value and threshold coefficient, we can get two thresholds. From Equation 15, we can calculate any inter-frame difference as well as the mean value of it in one window. It is shown in equation 16

$$u = \frac{\sum_{1}^{l-1} D_{i,i+1}}{(l-1)} \tag{16}$$

In equation 16, 1 is the size of sliding window, whose mean value approximately reflects the average value of neighboring frame difference in the whole shot; and with increase of window length, the value of approximation becomes higher. In light of abrupt and gradual transition s of shots, as well as differentiation between adjoining inter-frame difference in shot switching edge and that within shot, we calculate two adaptive threshold values. It is shown in equation 17-18.

$$Th = \lambda_1 \times u \tag{17}$$

$$Tl = \lambda_2 \times u \tag{18}$$

5.3. Description of the Algorithm

To implement the correct detection of video shot edges and reach expected effects, we should improve the algorithm's recall ratio and precision rate, which is the primary solution, together with reduction of computational workload for quicker detection speed. The proposed algorithm based on adaptive dual threshold is described as follows:

Input: complete volleyball video clip data;

Output: collection of abrupt and gradual transition shots

Make sliding window w, whose length is 1 and is initialized to 1=1. Set loop variable i=3, j=0 (sign of start frame with potential gradual transition), 2<i<n, where n is total

number of video frames; calculate inter-frame difference between the third and second frame and use it as the initial value of inter-frame difference mean. $u = D_{2,3}$

Step 1: add one-frame image in the follow-up frames to the window; use equation 15 to calculate the inter-frame difference $D_{i,i+1}$ between the new and previous frame; as per equation 16, compute the mean value u of adjacent inter-frame differences in the window; then with formula 17-18 and u, get high threshold Th and low threshold Tl;

Step 2: if i=n-1, it loops in the end frame; the detection is over and algorithm exits;

Step 3: compare the inter-frame difference between the new and nearby frames and the threshold; if $D_{i,i+1}$ >Th, it's Abrupt transition shot edge; mark the i frame as Abrupt transition frame and turn back to Step 1; if $D_{i,i+1}$ <Tl, it's within shot; with expression 16, calculate the mean value u of adjacent inter-frame difference in the window and return to Step 1;

Step 4: if j=0, mark j=i as the start frame with potential Abrupt transition and record the color histogram information of this frame; return to Step 1 or go to Step 5;

Step 5: utilize equation 15 to calculate the spaced inter-frame difference $D_{i,i+1}$ between the current frame and start frame with gradual transition ; compare it with Th; if $D_{i,i+1}$ >Th, the current frame is start frame with gradual transition and turn to Step 1; otherwise, return to Step 1.

6. Experiment Design and Discussion

The experiment employs Visual C++ 6.0. For better detection results, we chose several fragments from Volleyball match videos in 2008 Beijing Olympics as testing data. Since they're very important videos, the noise interference is relatively low. Also as mentioned before, sport video shot cuts are visually gentle. We make two threshold coefficient value $\lambda_1 = 13.5$ and $\lambda_2 = 2.8$.

In the experiment, we take the first kind of evaluation indicators: recall and precision ratio to measure the effect of shot detection algorithm. On the basis of ensuring high recall ratio, we try to enhance the accuracy rate of detection. The value of threshold coefficient is dependent on them as well.

Those video clips have different play durations. The number of shots and quantity of abrupt and gradual transition shots are not the same. So, experimental findings can prove as a whole the performance of the discussed algorithm. Besides, we compared it with general histogram difference algorithms. The experimental results are shown in Table 1-4.

Table 1. The Results of Abrupt Shot Boundary Detection Based on Histogram

Video clips	Mutations number	Detected number	False number	Missing number	Recall (%)	Precision (%)
Clip1	53	49	1	5	90.5	97.3
Clip2	122	115	4	11	90.9	96.5
Clip3	176	168	3	11	93.7	98.2
Comprehensive	351	332	8	27	92.3	97.6

Table 2. The Results of Abrupt Shot Boundary Detection based on Adaptive Dual-Threshold

Video clips	Mutations number	Detected number	False number	Missing number	Recall (%)	Precision (%)
Clip1	53	51	1	3	94.3	98
Clip2	122	120	0	2	98.3	100
Clip3	176	173	1	4	97.7	99.4
Comprehensive	351	344	2	9	97.4	99.4

Table 3. The Results of Gradual Shot Boundary Detection Based on Histogram

Video clips	Mutations number	Detected number	False number	Missing number	Recall (%)	Precision (%)
Clip1	6	4	1	3	50	75
Clip2	11	10	2	3	73.2	80
Clip3	19	17	2	4	78.9	88
Comprehensive	36	31	5	10	72.2	83.9

From the above Table 1-4, we see that the algorithm realized good effect for the detection of abrupt transition shots, with higher recall and accuracy rate. Also it reached satisfactory result in detecting shots with gradual transition, whose precision rate is low. But the detection algorithm based on histogram has low recall ratio and precision rate for the detection of both abrupt and gradual transition shots. Specifically, in the detection of gradual transition shots, the proposed algorithm performs much better than histogram-based algorithms on part of both recall and precision.

Table 4. The Results of Gradual Shot Boundary Detection Based on Adaptive Dual-Threshold

Video clips	Mutations number	Detected number	False number	Missing number	Recall (%)	Precision (%)
Clip1	6	6	1	1	83.3	83.3
Clip2	11	12	2	1	90.9	83.3
Clip3	19	21	3	1	94.7	85.7
Comprehensive	36	39	6	3	91.7	84.6

After analyzing videos, we found when it's close-up shot, people or objects are passing over quickly, it's considered the wipe change of gradual transition, which thus leads to wrong detection of gradual transition shots. When the camera shakes fiercely, the scene changes sharply, causing wrong detection of abrupt transition shots. For edges of abrupt transitions shots, color histograms of anterior and posterior frame of edges are much alike, which results in missing detection. Regarding the first two detections, we have to study further in the following works to reduce false detection of abrupt and gradual transition shots. Regarding the third detection, we consider splitting frame image to a few tiny sub-locks and compare them in two according frames. Unfortunately the computational amount will increase a lot. In whatever cases, our proposed algorithm proves simple, effective and applicable.

7. Conclusion

This paper introduces the current commonly on video shot boundary detection algorithm, and also introduces some mature techniques in these methods. According to the characteristics of televised volleyball video, this paper proposed the volleyball video shot adaptive double threshold detection algorithm based on feature, the algorithm uses video shot transform in vision, by calculating the color histogram distance as the absolute value of difference between frames in the HSV color model, then compare and judge the dual threshold, video shot boundary. The experiment shows that the detection method of shot boundary detection effect is good, in the detection of different video segments, recall and accuracy are relatively high, in addition to the current gradient detection technology is not mature, other tests have reached more than 90%, has good robustness.

References

[1] Z. Yuzhen, W. Jianyu and D. Yuewei, "Soccer video and adaptive dual threshold segmentation based on dominant color percentage", Journal of Nanjing University of Science and Technology (NATURAL SCIENCE EDITION), vol. 4, **(2009)**, pp. 432-437.
[2] Z. Yuzhen, Y. Ming, W. Jianyu and D. Yuewei, "The segmentation of motion compensation and adaptive dual threshold based shot", Computer science, vol. 9, **(2010)**, pp. 282-286.
[3] J. Yufu, W. D. Hong and H. Shenghong, "Research on segmentation algorithm of soccer video Adaptive double threshold based on microcomputer and application", vol. 19, **(2011)**, pp. 41-43.
[4] H. Tao, "Research and Simulation on the key technology of content based video retrieval", Wuhan University of Technology, **(2014)**.
[5] F. lulu, "The contents of video retrieval and classification method based on. Shenyang University", **(2013)**.
[6] Z. Shunhong, where Jiannong, "SUSAN algorithm based on double threshold", Computer Engineering, vol. 3, **(2012)**, pp. 206-208.
[7] J. hong and Z. Yuanhua, "Content retrieval and video processing technology based on computer", Journal of image and graphics, vol. 5, no. 4, **(2008)**, pp. 276-283.
[8] J. Haitao and H. A. Salam, "Scene Change Detection Techniques for Video Database Systems", Multimedia Systems, no. 6, **(1998)**, pp. 186-195.
[9] M. Toller, S. Lewis and M. S. Nixon, "Video Segmentation Using Combined Cues", Proc. SPIE, vol. 3312, **(1997)**, pp. 414-425.
[10] N. V. Patel and I. K. Sethi, "Video Shot Detection and Characterization for Video Databases", Pattern Recognition, vol. 30, no. 4, **(1997)**, pp. 583-592.
[11] H. J. Zhang, J. Wu and D. Zhong, "An Integrated System for Content-Based Video Retrieval and Browsing", Pattern Recognition, vol. 30, no. 4, **(1997)**, pp. 643-657.
[12] W. A. C. Fernando, C. N. Canagarajah and D. R. Bull, "Automatic Detection of Fade-In and Fade-Out in Video Sequences", IEEE International Symposium on Circuits and Systems, vol. 4, no. 6, **(1999)**, pp. 255-258.
[13] R. C. Gonzalez, R. E. Woods, R. Qiuqi and R. Yuzhi, "Digital image processing (Second Edition). Beijing: Electronic Industry Press", **(2008)**, pp. 228-238.
[14] Z. Yujin, "The content of the video information retrieval", Beijing: Science Press, **(2009)**.
[15] A. Hanjalic, "Shot Boundary Detection: Unraveled and Resolved", IEEE Trans. on Circuits and System for Video technology, vol. 12, no. 2, **(2002)**, pp. 90-105.

Author

Wei Liu. He received his B.S degree from Shenyang Sport University in 2001, and received his M.S degree from Northeast Normal University in 2006. He is a lecturer in Physical Educational Department of Heilongjiang University. His research interests include Information-based Physical Education, Sport sociology.

