International Journal of Multimedia and Ubiquitous Engineering
Vol.12, No.1 (2017), pp.281-290
http://dx.doi.org/10.14257/ijmue.2017.12.1.24

Research on Video Game Scene Annotation in Basketball Video

Changliang Zheng[1], Zhiqian Zhang[2], Zhaoxin Liu[3], Hepeng Zang[2], Hongli Huang[2] and Jie Ren[4,*]

[1]Changshu Institute of Technology, Changshu 215500, China,
[2]Harbin Engineering University, Harbin 150001, China,
[3]Harbin Vocational College of Science and Technology, Harbin 150001, China,
[4]Harbin Institute of Physical Education, Harbin 150001, China
12701069@qq.com
[*]Corresponding author

Abstract

This paper is mainly researching on the basketball match video annotation. In the process of annotation, shot is the process unit, which boundary detection has high priority. This paper detects the shot boundary based on the difference between the histogram of key region. In the contracture of the sports video, scene's level is higher than shot's and it can be obtained by clustering shot into different classes. RFCM cluster can be used to classify the shots and annotate the scene.

Keywords: *Sports Video, Video annotation, Tactics annotation*

1. Introduction

Video annotation is done manually by the human, and then used the method of text retrieval [1-10]. This approach is inefficient and heavy workload, in the case of a sharp expansion of the current amount of data, the method is no longer desirable [11-20].

To solve this problem, in the last century ninety's, content-based video retrieval came into being. More accurate point [21-27], the content based video retrieval is supposed to be based on sample retrieval, which is based on the sample [28-34].

It extracts the query samples and the underlying visual features, audio and text information from the query samples. However, the existence of "semantic gap" between low-level features and high-level semantic concepts hinders the development of this method. Currently, how to effectively get video content [35-40], so that it becomes easy to search and easy to interact with the data has become the core content of video annotation [41-43].

In view of features of basketball match videos, the scene of such video can be divided into three kinds: close-up, in-play, and penalty shot [44-50]. They can stand for most contents of basketball match videos. Here in marking shooting semantic event in the following part, we make detailed classification of scenario in accordance to needs [51-58].

In order to make annotation of scene, we firstly do shot boundary detection of basketball match video. The video is separated into many single shots; then we fetch shot characters from therein [59-61]. Shot boundary detection is foundation to scene annotation and also the basis to mark high-level semantic annotation of basketball match video in follow-up work.

Scene is video structure unit composed of shots which are however higher than shots. Scene construction can be implemented through shot categorization [62-63].

ISSN: 1975-0080 IJMUE

2. Shot Boundary Detection

2.1. Choice of Color Space

Color is human eyes' different feelings about light of different frequencies. Color is associated with light of different frequencies which exist subjectively; but also color is objectively sensed. That's why there's cognitional difference. Human's awareness of color went through a long process, till modern times, the cognition accomplished. However till today, we can't say human has understood fully color and can express it accurately. The term "color space" originated from the West, renamed "color gamut". In chromatics, people would use one dimension, two dimension, three dimension even four dimension spatial coordinate to describe one color and create color model. The color scope defined in the way of coordinate system is called color space. Commonly used color space RGB, HSI, YCbCr and HSL, *etc.*

2.1.1. RGB Color Space

At present, a majority of computer display system adopts RGB color space. The space uses the level of red, green and blue to define color. The coordinate axis R, G, B represents respectively color red, green and blue. Each color component is divided into 256 units. The value of each color component at origin is 0 and corresponding color is black. The value of three color component of vertex which is diagonal to the origin is all 255 and according color is white. The point of which the three color component has same value is gray-scale pixel point. All gray-level pixel points are on the diagonal line of color space cube, which is gray-level straight line. Apart from two vertexes whose color is respectively white and color and the vertex on the coordinate, other vertices on the color cube are corresponding to cyan, yellow and purple.

2.1.2. HSI Color Space

HSI color space is described in hue, saturation and illumination. Hue means pure color attribute; it can be measured by angle -180-180 or 0-360 degree. Saturation means deep or light degree of color; for instance red, which can be dark red or light red because of different concentration. The concentration can be scaled in percentage between 0-100%. Illumination is disjunct with color information. It means bright or dark degree of color between 0-100%, meaning from black to white. The pixel point getting closer and closer to central vertical axis suggests the color is much more unsaturated. The conversion from color space to color space is formulated as follows.

$$\begin{cases} \theta = \cos^{-1}\left\{ \dfrac{\frac{1}{2}[(R-G)+(R-B)}{\sqrt{(R-G)^2+(R-B)(G-B)}} \right. \\ H = \begin{cases} \theta & B \le G \\ 360-\theta & B \ge G \end{cases} \\ S = 1-\dfrac{3}{(R+G+B)}[\min(R,G,B)] \\ I = \dfrac{(R+G+B)}{3} \end{cases} \tag{1}$$

2.1.3. YCBCRF Color Space

In YCBCRF color space, Y is brightness; Cb & Cr is chroma, where Cb is blue component and Cr is red component. In JPEG standard, RGB image is converted to YCBCR color space. YCBCR color model is applied in CCIR601 coding scheme. The sampling ratio of three components is Y: Cb: Cr=4:2:2. That ratio is adopted because human eyes are more susceptible to brightness information change than chromatic information. Formula 2 shows the conversion from color space to YCBCR color space.

$$\begin{bmatrix} Y \\ C_r \\ C_b \end{bmatrix} = \begin{bmatrix} 0.254 & 0.503 & 0.089 \\ -0.156 & -0.259 & 0.487 \\ 0.467 & -0.356 & -0.061 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 17 \\ 128 \\ 128 \end{bmatrix} \quad (2)$$

2.1.4. HSL Color Space

HSL color space is one of the most widely used color space, where H refers to hue; S means saturation; L stands for luminance. HSL color space can help get diversified colors through color channel change and overlay of hue, saturation and luminance. Those colors cover almost all colors which can be perceived by human vision.

The algorithm used to convert RBG color space to HSL color space is described below:

Setp1: R, G, B value normalized to [0,1].

Setp2: Take out R, G, B in the maximum and minimum value is divided into maxcolor and mincolor

Setp3: L=(maxcolor +mincolor)/2

Setp4: If the maximum and minimum values are the same, that is, the S is defined as 0, while the H is not defined, usually written in 0.

Setp5: Otherwise, judge L:

If L<0.5,S=(maxcolor-mincolor)/(maxcolor+mincolor)

If L>=0.5,S=(maxcolor-mincolor)/(2.0-maxcolor-mincolor)

Setp6: If R=maxcolor,H=(G-B)/(maxcolor-mincolor)

If G=maxcolor,H=2.0+(B-R)/(maxcolor-mincolor)

If B=maxcolor, H=4.0+(R-G)/(maxcolor-mincolor)

Setp7: H=H*70.0.

According to the characteristics of the basketball game, this paper chooses to use the value of color space to detect the boundary of the shot.

2.2. Color Histogram of Key Areas

The basic idea of inter-frame difference method based on color histogram is: for adjacent two-frame image, their color histograms are respectively acquired; make difference between the two histograms; if the result is bigger than certain threshold, it’s believed there is shot cut.

Basketball match broadcasting has unique features, so that the video has its own characteristics, that is, the area of play court in each frame image is basically stable. With that feature, by combining the use of “golden cut rule” in shooting field, we chose color histogram inter-frame difference method based on key area to detect boundaries of shot and that the computing efficiency is enhanced.

2.2.1. Few Kinds of Common Color Histograms

Common color histograms include global color histogram, cumulative color histogram and dominant hue histogram.

(1) Global Color Histogram

The global color histogram describes color composition and distribution of the whole image, *i.e.* what colors are included in the image and how many times they're used there. Global color histogram is not sensitive to geometric transformation like not big image translation and scaling and rotation of which observation axis is axle center, nor is sensitive to variations of image quality. Hence, global color histogram is usually applied to measure differences between two images in color's global distribution.

(2) Cumulative Color Histogram

In some images, their features can't be set all available values. At this moment, global histogram can have some zero values. Those zero values can affect calculation of similarity measurement, as a result, similarity measurement can't properly reflect color differences between two images. The problem can be avoided to a certain extent by expanding the value range of image feature to reduce feature value. However, there's still a question. We can take H component for instance. To quantify two similar colors, it may increase or decrease the distance between them; besides, similarity distribution of tonal feature is not even, *e.g.* the distribution range of yellow's similar colors is narrower than sparse color, so simple uniformly-spaced quantification can't reflect accurately the hue difference between two images. To solve the histogram zero-value problem mentioned above, we can apply cumulative color histogram method.

The cumulative histogram of the image is also a one dimensional discrete function:

$$I(k) = \sum \frac{n_k}{N}, k = 0,1,...L-1 \tag{3}$$

(3) Dominant Hue Histogram

In an image, usually very few kinds of colors can contain most pixels of the image. Different color appears there at different frequency. In this case, through statistics, we can get several colors which occurs the most frequently and use them as major color. Since colors which appear infrequently in color histogram is generally not main content of an image, color matching effect won't be impaired for the use of dominant color.

2.2.2. Color Histogram of Key Area

In photography art area, there's a photography rule named golden cut. The method cuts image by 3:5:3 in axis X and Y direction. The area where four lines intersect is the theme that the image represents. In view of features of basketball match video, we use golden cut rule to make regional division of video frame image, choosing the central region for processing. Figure 1 presents pictures of match and close-up images which are treated by golden cut rule. Based on our statistical observations of abundant video frames, we think images in R2 region can be used to reflect enough the content of the whole image.

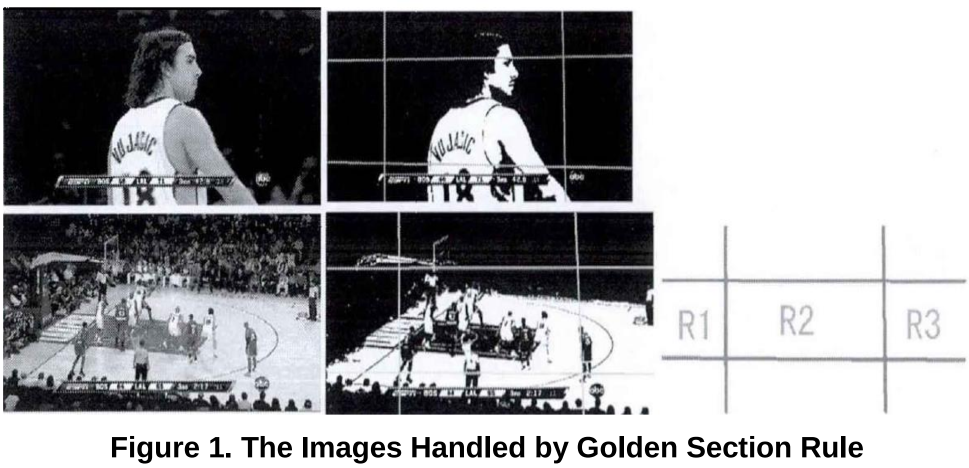

Figure 1. The Images Handled by Golden Section Rule

In doing shot boundary detection, we can use color histogram of R2 region of basketball match video frames to calculate inter-frame difference. If it's bigger than threshold value, it's believed that shot changed.

2.3. Video Shot Categorization

Clustering is a mathematical process of classifying data set per the internal structure or specific pattern of such set. In data set, element of each class show certain similarity, which is one of the significant bases for classification. In the framework of fuzzy clustering, each feature vector belongs to one or more clusters simultaneously with the degree of membership varying between 0 and 1. Hence, the problem of data classification can be converted to that of quantifying similarity or dissimilarity degree between objects. In fuzzy relation clustering, the problem of data clustering can be changed to that of relationship between similarity and dissimilarity degree. RFCM's actually a kind of fuzzy relation clustering algorithm.

2.3.1. Merits and Shortcomings of Fuzzy Clustering Algorithm

FCM is a widely used common clustering technique. It's simple and can effectively find the minimum objective function J_m .Given data set $X = \{x_1, x_2, ...x_n\}$, In order to gather the data set C into the X class, the J_m is defined:

$$\min_{(U,V)} \left\{ J_m(U, v) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m D_{ik}^2 \right. \tag{4}$$

However FCM has shortcomings such as huge computational work, too sensitive to noises; too bigger or smaller elements in abnormal data would affect the mean value of calculation; it can perform data clustering only for those distributed in the form of convex set.

2.3.2. Comparison of RFCM and FCM.

(1) RFCM is extended application of FCM

When feature space is of high dimension, and in data collection, the number of data elements is very few or lots of feature values are lost, only the relation between data can be measured. At this moment, FCM is too weak to deal with such data; instead, RFCM is utilized for classifying. It's seen RFCM gains wider application than FCM.

(2) Improvement of computing efficiency

Whether relational data can be availed depends on the pairwise distance between elements. RFCM is workable for the problem that requires good computing efficiency. One of its merits is although element data is not directly used, properties evaluated by it can be shared by all elements. Another merit is RFCM automatically inherits FCM's excellent property of data convergence.

Set matrix $R = [r_{ij}]$ is relational matrix determined by pairwise distance. Unlike FCM, it is objective function is defined as $K_m(U)$.

$$K_m(U) = \sum_{i=1}^{m} (\sum_{j=1}^{n} \sum_{k=1}^{n} (u_{ij}^m u_{ik}^m \delta_{jk}^2) / (2 \sum_{t=1}^{n} u_{it}^m)) \tag{5}$$

Furthermore, most experiments revealed that in practical use, when feature dimension of data is too big, with RFCM, it requires fewer calculation times than FCM in each iteration.

2.4. Basketball Match Video Shot Clustering Application

To realize scene classification and annotation, it's necessary to fetch key frames in a shot, with color histogram of key area as feature of subsequent clustering. Here we employ adaptive K-mean clustering method to extract key frames in a shot. In basketball competition video, frame image in a shot contains similar information and color histogram. We pick up intermediate frame from the acquired key frame sequence as shot's key frame.

The HOG similarity of histogram between two frame images is defined like:

$$s(p,q)=\sum_{u=1}^{B}\min\{p^{(u)},q^{(u)}\} \tag{6}$$

Among them, p and q are two histograms, if they are similar, S=1, and the dissimilarity is defined as d=1-S. Therefore, the dissimilarity matrix of N image HOG is calculated.

$$D=[d_{ij}]_N=\begin{bmatrix} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{21} & \dots & d_{2N} \\ \dots & \dots. & \dots & \\ d_{N1} & d_{N1} & \dots & d_{NN} \end{bmatrix} \tag{7}$$

Where the diagonal elements d_{ii}=0, other elements of the value of d_{ii} for the difference value the i frame image and the j frame image.

Since HOG dissimilarity is fuzzy, we select RFCM algorithm for clustering.

3. Experimental Analysis and Results

3.1. Environment of Simulation Experiment

Hardware environment is Inter(R) Pentinum® 3.0GHZCPU, 2GM memory. Software environment is WINXP OS, Visual C++6.0, Visual c#2008, OPenCV, Aforge.net.

3.2. Experimental Results and Analysis of Shot Boundary Detection

The video data for experiment is extracted from NBA 2014 All-star game. The first video lasts 20 minutes, including 140 shots; the second video lasts 40 minutes with 260 shots. It is shown in Table1.

Table 1. The Results of Shots Boundary Detection

	Number of people split	Number of independent split lenses	Correct number of detected	Number leakage	Number of false check	Precision	Recall
Video1	146	165	125	17	25	84.9%	89.7%
Video2	256	265	176	43	36	85.6%	83.8%

Where,

$$\text{Recall}=\frac{\text{Correct number of detected}}{(\text{Correct number of detected}+\text{Number leakage})}$$

$$\text{Precision}=\frac{\text{Correct number of detected}}{(\text{Correct number of detected}+\text{Number of false check})} \tag{8}$$

The experiment shows that the shot boundary detection method based on histogram of key area achieves good result.

3.3. Experimental Results and Analysis of Video Shot Classification

The video data for experiment is extracted from NBA 2014 All-star game. The first video lasts 30minutes. It is shown in Table2.

Also it shows that the proposed method realizes good effect in basketball match video shot classification. As seen from data, the algorithm here is more effective to close-up than in-play and penalty shot. That's because frame images of some in-play clips have similar key area color histogram with frame images of penalty shot. How to distinguish them accurately will be our research concern in the future work.

Table 2. The Results of Shots Classification

	Hand marking	Auto marking	Correct classifica tion	Numbe r leakage	Number of false check	Precision	Recall
Match play	88	79	56	23	26	74.3%	69.2%
Close-up	103	106	79	29	29	76.6%	78.2%
Free throw	19	19	12	13	9	69.3%	75.3%

4. Conclusion

In this paper, with the characteristics of the key region of the color histogram, shot boundary inspection of basketball video. It is measured by RFCM clustering algorithm, the segmentation of the scene classification, in order to achieve the basketball video game, the purpose of scene annotation. The experiment proved that this method has better effect of scene annotation.

References

[1] Z. Nan, "Video shot segmentation and its application in video retrieval", Southwestern University, **(2009)**.
[2] S. Yanran, "Content based image retrieval and video annotation", Fudan University, **(2009)**.
[3] B. Hongfei, "Scene image text extraction method research and application", Fudan University, **(2009)**.
[4] Q. Xueming, "Research on Key Technologies of semantic based video browsing system", Xi'an Jiao Tong University, **(2007)**.
[5] H. T. Wen, "Modern basketball pick and roll with performance analysis and practice research", Beijing Sport University, **(2012)**.
[6] Y. Lelin, "Research and analysis of video semantic retrieval based on visual information", Beijing University of Posts and Telecommunications, **(2012)**.
[7] N. Revitalization, "Research on topic modeling and content analysis of soccer video", Xi'an Electronic and Science University, **(2012)**.
[8] X. Ruxi, "Study on target tracking algorithm in complex environment", Chongqing University, **(2012)**.
[9] K. Jia, "Research on semantic based video event detection and analysis", Jiangsu University, **(2013)**.

[10] G. Guangyu, “Research on video structure analysis of cosmos and automatic cataloguing technology”, Beijing University of Posts and Telecommunications, **(2013)**.
[11] W. Xinping, “Research on the visual search strategy and prediction ability of the different roles of female volleyball players and its neural mechanism”, Shanghai Institute of Physical Education, **(2010)**.
[12] L. Xun, “Basketball object segmentation and tracking technology research”, Beijing University of Technology, **(2005)**.
[13] C. Shiju, “Key technology research of video analysis in basketball games”, Beijing University of Technology, **(2006)**.
[14] L. Huifang, “Multi level analysis and retrieval of tennis video”, Nanjing University of Science and Technology, **(2009)**.
[15] L. Xiaowei, “Sports video analysis and personalized customization”, Wuhan University of Technology, **(2010)**.
[16] W. Ning, “Preliminary study on basketball video analysis system”, Qingdao University of Science & Technology, **(2013)**.
[17] L. Kai, “Feasibility study and design of real time tracking and analysis system for basketball video and tactics”, Capital Institute of Physical Education, **(2014)**.
[18] L. Yingying, “A study on the detection method of exciting events in soccer video based on HCRF”, Xi'an Electronic and Science University, **(2013)**.
[19] D. Liwei, “A study on the detection method of exciting events in soccer video based on HCRF model”, Xi'an Electronic and Science University, **(2014)**.
[20] F. C. Jian, “Study on mining method of video content”, Based on the structure of national defense science and Technology University, **(2008)**.
[21] Z. Guangyu, “Research on sports video content analysis method based on the information of players' behavior”, Harbin Institute of Technology, **(2009)**.
[22] W. Qian, “Research on basketball player behavior prediction”, Based on the video of Central South University, **(2012)**.
[23] X. Wenjuan, “Soccer video exciting scoring event detection”, Xi'an Electronic and Science University, **(2012)**.
[24] W. Yu, Z. Lizhu and X. Chunxiao, “The video semantic model and evaluation criteria”, Chinese Journal of computers, vol. 3, **(2007)**, pp. 337-351.
[25] S. yingchun, “History of content-based video retrieval research on semantic extraction based on”, Nanjing University of Science and Technology, **(2005)**.
[26] C. Jianyun, “Research on semantic content analysis technology of sports video”, University of national defense science and technology, **(2005)**.
[27] P. Liang, “Target detection and tracking in video of football match”, Nanjing University of Science and Technology, **(2006)**.
[28] L. Xuezhao, “Research and implementation of content based sports video description, management and browsing”, Graduate University of Chinese Academy of Sciences (Institute of computing technology, Institute of computing technology), **(2003)**.
[29] L. Xueying, “Semantic analysis of sports video based on rules”, Qingdao University of Science & Technology, **(2011)**.
[30] K. Changlang, “Research on video abstract technology for the game of basketball”, Zhejiang University, **(2015)**.
[31] Z. Nannan, “Research on automatic identification technology of players in sports video”, Beijing University of Posts and Telecommunications, **(2015)**.
[32] Y. Ming. “Application and development of sports performance analysis techniques in basketball games”, Beijing Sport University, **(2014)**.
[33] W. Hui, “Application Research of video annotation technology in sports competition analysis”, North China University of Technology, **(2011)**.
[34] S. Bin, “Mobile background under the research on moving object detection and tracking technology”, Electronic measurement and instrument of, vol. 3, **(2011)**, pp. 206-210.
[35] P. Yanfang, “Research on moving target detection and tracking algorithms”, Wuhan University of Technology, **(2010)**.
[36] T. Penghui, “Research on moving object detection and tracking in video image”, Chang'an University, **(2013)**.
[37] W. Yanhong, “Research on moving target detection and tracking algorithm based on OpenCV”, Hangzhou Dianzi University, **(2014)**.

[38] L. Xinhua and S. Zhongke, “Based on dynamic template matching of air moving target detection and tracking control”, Electronic measurement and instrument, vol. 10, **(2010)**, pp. 935-941.
[39] Y. Baohong, Z. Dexiang and Z. Lingjun, “Detection and tracking of video moving objects based on OpenCV”, Application of computer system, vol. 5, **(2013)**, pp. 90-93.
[40] Y. Wei, B. Xiaofang and C. Peiheng, “An improved moving object detection and tracking method”, TV technology, vol. 1, **(2014)**, pp. 180-182.
[41] F. Huiyuan, “Research on intelligent video scene understanding technology for crowd control”, Beijing University of Posts and Telecommunications, **(2014)**.
[42] F. Weiguo, “Moving object extraction and rapid retrieval of massive objects in video surveillance”, University of Science & Technology China, **(2015)**.
[43] T. Zhiwei, “Research and implementation of video structured description for public security business”, Shanghai University, **(2015)**.
[44] Q. Jianming, “Intelligent video surveillance in moving target detection and tracking technology research”, Xi'an Electronic and Science University, **(2015)**.
[45] H. X. Winter, “Research on access to video text based on feature fusion”, Beijing University of Posts and Telecommunications, **(2010)**.
[46] C. Huanhuan, “The image content representation and classification based on Bayesian network”, National defense science and Technology University, **(2011)**.
[47] J. Peng, “Research on Key Techniques of semantic based video content extraction and analysis”, Nanjing University of Aeronautics & Astronautics, **(2009)**.
[48] L. Haipeng, “Visual semantic representation of language and its application in automatic scene description system”, Beijing University of Posts and Telecommunications, **(2011)**.
[49] G. Guangyu, “Research on video structure analysis of cosmos and automatic cataloguing technology”, Beijing University of Posts and Telecommunications, **(2013)**.
[50] L. Heng, “Geographic information annotation technology and application of massive multimedia data”, University of Science & Technology China, **(2014)**.
[51] W. Longlong, “Film structure of video scene analysis and recognition technology research”, Harbin Institute of Technology, **(2014)**.
[52] T. Jinhui, “Research on some problems of video semantic annotation”, University of Science & Technology China, **(2008)**.
[53] G. Zhiwei, “Research on video retrieval for structured data”, University of Science & Technology China, **(2008)**.
[54] F. C. Jian, “Study on mining method of video content”, Based on the structure of national defense science and Technology University, **(2008)**.
[55] W. Min, “Research on the algorithm of teaching video annotation”, Xi'an Electronic and Science University, **(2014)**.
[56] L. Jing, “Study on behavior recognition fusion of multiple visual objects”, Beijing Institute of Technology, **(2015)**.
[57] L. Hongying, “Research on content based image and video retrieval technology and its system implementation”, Hunan Normal University, **(2006)**.
[58] G. Xiaochuan, L. Mingjie, W. Jinglu, D. G. Dong and Qianrong, “Frequent shot set of video scene segmentation method based on”, Computer applications and software, vol. 6, **(2011)**, ppp. 116-120.
[59] Liu and M. Yan, “Research of moving object detection and tracking algorithm”, Journal of Hebei United University (NATURAL SCIENCE EDITION), vol. 1, **(2015)**, pp. 65-70.
[60] T. Xiaofeng, L. Qingshan and L. Hanqing, “Analysis of sports video”, Journal of computer science, vol. 7, **(2008)**, pp. 1242-1251.
[61] J. Yu, “Research on the motion video tracking technology based on mean shift algorithm and color histogram algorithm”, Journal of Soochow University (Engineering Science Edition), vol. 2, **(2012)**, pp. 33-36.
[62] Q. Ping, S. G. Qu, K. Tao and Y. L. Zhao, “Based on underlying visual information of sports video intelligent analysis”, Journal of sports adult education, vol. 3, **(2012)**, pp. 49-51.
[63] X. Yan, “Research on moving target tracking and detection algorithm in Volleyball video”, Science and Technology Bulletin, vol. 7, **(2015)**, pp. 160-162.

Author

Chang liang Zheng, He received his B.S and B.S degree from Shenyang Sport University. He is a lecturer from Changshu Institute of Technology. He is in the research of Physical Education and Training