HAL open science

Development and Application of Tennis Match Video Retrieval Technology in Multimedia Education

Shehua Cao

To cite this version:

Shehua Cao. Development and Application of Tennis Match Video Retrieval Technology in Multimedia Education. 4th Conference on Computer and Computing Technologies in Agriculture (CCTA), Oct 2010, Nanchang, China. pp.192-197, 10.1007/978-3-642-18369-0_21 . hal-01564852

HAL Id: hal-01564852

https://inria.hal.science/hal-01564852v1

Submitted on 19 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Development and Application of Tennis Match Video Retrieval Technology in Multimedia Education

Shehua Cao[1]

School of P.E., East China Jiaotong University
Nanchang, Jiangxi ,P.R.China
Shehua Cao，*shehuacaonc@sina.com*

Abstract. In order to improve the application effect of tennis match video in multimedia education, this paper puts forward a kind of method for retrieving the spectacular Rally scenes in tennis match video and classified the shots into Global View Shots and Non-Global View Shots. The Support Vector Machine (SVM) is for identifying the important types of audio in shots, applying the audio information in shots and setting up corresponding rules for retrieval. The results show that applying the retrieved spectacular Rally scenes in tennis match video in tennis multimedia education is line with most students' psychological situation and greatly popular with students.

Key words: Tennis Match Video, Visual Features, Movement Features, Audio Features

1. Introduction

With the emergence of many tennis match videos, the tennis match video retrieval technology has become a study topic urgently and strongly needed in application. The tennis match is very long, the proportion of the Non-play Shots is large and generally, the audience are usually interested in the match scenes or the spectacular Rally scenes in match, therefore, automatically retrieving the spectacular scenes from the tennis match video which the audience is interested in can save the audience's time for appreciation of match and make the audience be able to enjoy the match's spectacular Rally scenes in limited time. In addition, the spectacular Rally scenes extracted can also be applied in tennis multimedia education, which is of great research significance and application value in development of the retrieval technology for the spectacular Rally scenes in tennis match video.

[1] Corresponding author. Tel.: +86-13507918839
*E-mail address:*shehuacaonc@sina.com

2. Structure of Tennis Match Video

The Rally shots, the players' close-up shots, the audiences' shots, the playback shots, the shot cuts and other shots which the audiences are interested in tennis match video are with the predicable sequential structure. For the Play Shots in tennis match video, the number of play targets is fixed and the size of play targets can hardly be changed; the color of court accounts for the overwhelming majority of the video and the main color ratio is relatively high. The majority of Play Shots are Global View Shots, however, the Non-play Shots including the players' close-up shots and the audiences' shots are generally regarded as the Non-Global View Shots. The video's major audio events include the sound of ball-striking, the hurrahs of audiences, the sound of applause and the sound of the commentators. These sounds sometimes mix with each other, which even mix with the noise from the auditorium. But the audio information and the match proceeding state are closely related, like the ball-striking sound shows the Play state and after the Rally shots, there often come the sound of applause. It is thus evident that the retrieval of some audio events can help us analyze the tennis match video in a better manner.

Through the analysis on the structure of tennis match video, we can find that the shot boundary detection is more suitable for the retrieval of spectacular Rally scenes in tennis match than the audio segmentation. For this reason, in the retrieval, we should firstly measure the similarity of shots based on the features of movement and major colors and design a two-category classifier, which classified the shots into the Global View Shots and Non-Global View Shots; and then apply the SVM for indentifying the meaningful audio events (like the sound of ball-striking and applause); apply the visual and audio features to set up the rules for heuristic procedures, detect the issues which the audiences are interested in the match video; finally set the spectacular degree model of the Rally scenes and arrange the spectacular degree in the from-high-to-low order.

3. Classification of Tennis Match Video Shots

Due to the fact that the majority of Play Shots in tennis match video are Global View Shots, generally regarding the Serve scene as the Rally scene is not suitable[1], however, when classifying the shots into the Global View Shots and the Non-Global View Shots, we are not required assuming the color of court in advance, which is thus more suitable for the tennis match video shots. Apart from that, the classification of shots is not only required considering the similarity of low-level visual features, but people's perception and attention on movement. It should suit the outdoor tennis match video with relatively apparent changes of bright.

3.1 Classification Based on Features of Major Colors

The tennis match video mainly includes four typical types of shots, namely, the Global View Shots, the Medium View Shots, the Close-up View Shots and the

Audience View Shots. All Global View Shots have some common features of colors. The color of court in the Global View Shots accounts for the majority of the picture frame, making the main color ratio relatively high; apart from that, when comparing with the Non-Global View Shots, the colors of Global View Shots also distribute evenly. Therefore, the tennis match video can define the features of colors in Global View Shots through the description of key frames and the spatial distribution of major colors. We adopt the HSV color solid as the color models and classify the shots through comparing the key frames of each shot and referring to the similarity of key frames[2].

3.2 Classification Based on Features of Movement

The tennis match video always follows the attention-focus area and places it in the central part of the picture. Generally speaking, the attention-focus movement area and the non-attention-focus movement area have two differences in the spatial and time domain; in addition, the more intensified movement events are usually more attractive to audiences than the slow movement events[3]. Therefore, the conclusion of movement features should be based on the attention-focus movement model and the number and size of attention-focus movement area in the picture frame should be extracted according to the features and rules of tennis match[4]. When watching the tennis match, people's attention is usually focused on the players and the ball in the match shots, while when appreciating the close-up shots of players, people are interested in the players' some interesting behaviors. Through careful analysis, we have concluded two movement features of tennis match video: namely, (1) there are two attention-focus movement areas in the match shots (the players and tennis' movement areas); the attention-focus movement areas in close-up shots are usually some player's behaviors; for the audience shots, the number of attention-focus areas is uncertain usually. (2) The sizes of players and balls are seldom changed; while the sizes of the attention-focus movement areas in other Non-play Shots are not fixed.

After the color and movement features of key frames are ascertained, we measure the similarity of two shots according to each key frame's movement and main color features and classify the shots into the Global View Shots and the Non-Global View Shots. When comparing with the Non-Global View Shots, the color features of Global View Shots include more colors of play court. Upon defining the main color ratio as the percentage of the main color pixel to the corresponding picture frame's all pixels, we conclude the main color ratio of the Global View Shots is higher.

4. Audio Identification in Tennis Match Video

4.1 Audio Type Identification Based on SVM[5] [6]

Although most Global View Shots in tennis match video are Play Shots and the majority of Non-Global View Shots are Non-play Shots, we can not only depend on the visual features to correctly identify the Play Shots in tennis match video. The typical tennis match video has the special audio features. Applying the audio information of shots into the classifications of shots can detect the Play Shots in a more satisfied manner. The SVM boasts many advantages which is easy for training, needs less samples in training and can give consideration to the inaccuracies in study and the capacity of generalization. Therefore, the SVM is suitable for identifying the types of audio in tennis match video shots.

The tennis match video's audio events include the sound of ball-striking, the sound of applause, the hurrahs of audiences, the sound of the commentators, the sound of referee and other unpredictable sounds from the site. During the whole play process, the commentator is almost commenting the match with a placid sound, even the match becomes brilliant, the commentator seldom sounds emotionally. For this reason, in tennis match video, there are not so much detection information for the sound of the commentator's comment. Furthermore, in tennis match, after each Rally shot, there will usually come the sound of applause. Comparatively speaking, the ratio of hurrahs is relatively low. Therefore, the sound of ball-striking and applause are the most significant audio events in the Rally match video, which can be directly used in differentiating Play Shots and Non-play Shots. For the reason that the sound of ball-striking and applause will not come at the same time, when applying the SVM in indentifying the audio events of the sound of ball-striking and applause, we should separately train a SVM for each audio part, namely, we could train the SVM with two kinds of audio which are separately for identifying the sound of ball-striking and applause.

4.2 Detection of Interested Events in Tennis Match Video

In tennis match, the players can get points from counter attacks. The audiences are often interested in Rally shots. Apart from that, serving a ball is also an important way for players to get points. Therefore, in addition to the Rally shots, the serve shots are the events which the audiences are further interested in. The RePlay Shots are often the repeatedly played brilliant scenes in the match, which can be used in retrieval for spectacular scenes. For this reason, the detection of interested events covers the Rally shots, the Ace, the Missed Serve and the RePlay Shots. We can detect the shots in tennis match video according to the visual and audio features of these typical shots and set some heuristic procedures for detection of interested events, for instance, apply the differences of the number of ball-strikes in the shots to differentiate the Rally shots and the Serve shots; differentiate the Ace and the Missed Serve in the serve shots through identifying whether there is he sound of applause at the end of the

shots; and differentiate the Play Shots and the Break Shots through identifying whether there is the sound of ball-striking in shots. Four rules can be worked out for the detection of interested events in the tennis match video shots:

Rule 1: IF there is no sound of ball striking in one Global View Shot and even there is no sound of applause at the end of such shot, THEN this shot belongs to a Break Shot, ELSE it is a Play Shot.

Rule 2: IF there is not only the sound of ball striking in one Global View Shot and even there is the sound of applause at the end of such shot, THEN this shot belongs to a Play Shot, ELSE it is a Break Shot.

Rule 3: IF there is no sound of ball striking in one Global View Shot and even there is no sound of applause at the end of such shot, THEN this shot belongs to a Missed Serve Shot, ELSE judge whether the number of ball-striking sound is 1; if the number of ball-striking sound is 1, THEN this shot belongs to an Ace Shot, ELSE it is an Rally Shot.

Rule 4: IF there is gradual change before and after the Break Shot, THEN this shot belongs to a Replay Shot, ELSE it is not a Replay Shot.

4.3 Retrieval and Sorting of Spectacular Rally Scenes in Tennis Match Video

The Rally shots in tennis match video are the Play Shots and are also the shots which the audiences are interested at. When there's no shot cut in the Rally scenes, the Rally scene is made up of one Rally shot; therefore, the spectacular Rally scene is actually one spectacular Rally shot. The retrieval for the spectacular Rally scene is retrieving the brilliant shots from the Rally shots detected from the above section. In order to comment on the spectacular degree of the Rally shots, we set the spectacular degree model for the Rally shots according to the features and rules of the tennis match and the judge on the spectacular degree of the Rally shots by most people. When watching the match in tennis education, the students are usually more eager to see the long-time counter attach scenes. Therefore, the lasting time of one brilliant Rally shot is one of the important factors for evaluating the spectacular degree of the shot. Moreover, after one great Rally shot, the audiences usually give more enthusiastic applause. We can define the Rally spectacular degree model based on this law.

Generally, after one great Rally shot, the tennis match video will Replay the brilliant parts of this shot from different visual angles. Therefore, the Rally live broadcast shot which is just after the Replay scene is one spectacular Rally shot, and the Replayed Rally shots are usually the most brilliant shots in the whole match video, which should be arranged in the forefront according to the from-high-to-low order of spectacular degree. The retrieval methods for the spectacular Rally scenes are as follows:

(1) Judge whether the current Rally shot in the set of Rally shots is Replayed. If the answer is positive, mark this shot as RRally shot;

(2) Take all RRally shots marked as spectacular Rally shots;

(3) Calculate the RH values of all RRally shots marked and the Rally shots unmarked;

(4) Set the threshold value $T_{Rally,}$ *if the RH of some unmarked Rally shots meet the standard of* $RH \geqq T_{Rally}$*, then this shot should be marked as spectacular Rally shot;*

(5) Arrange all the marked RRally shots and Rally shots according to the from-high-to-low order of the RH values, and arrange all RRally shots before the Rally shots.

5. Application Effect of Spectacular Rally Scenes in Tennis Match Video

In order to test the performance of the retrieval method for the spectacular Rally scenes in the tennis match video, we applied the same in the tennis match multimedia education in East China Jiaotong University(ECJTU). All the students having the lesson are the 2007-grade public sports tennis optional course students, who generally expressed that the video effect is very good and beneficial for learning the skills of tennis during the lessons. We sampled 90 students from the total of 260 students watching the tested scenes as the test targets and required them scoring the spectacular degree for the great Rally scenes, with the total score of 100. They respectively scored for the said scenes, making us receive the average score of 84.35, which is relatively satisfied.

6. Conclusion

(1)The computational complexity obtained only through applying the visual features and adopting the hierarchical clustering algorithm in classifying the video shots is fairly considerable, which is not suitable for the retrieval of tennis match video. Applying the SVM audio identification technology and integrating the movement and visual features in video retrieval is convenient for extracting the size and number of movement areas in picture frame which people are interested at, as well as helpful for classifying the shots in tennis match video and enhance the application effect of the same.

(2)Automatically extracting the reference key frames of Global View Shots based on the color and movement features and applying the similarity of the key frames and reference key frames of shots in classifying shots do not require the prior knowledge of the color of court. Such classification of shots is especially suitable for the tennis match video with the color of court not ascertained and the changes of bright of video relatively obvious.

(3)When applying the SVM in identifying the important types of audio, we worked out four heuristic procedures for detection the interested events, which can improve the detection precision. The high identification rate of Rally events is extremely important for the retrieval of spectacular Rally scenes in the tennis match video.

(4)The visual retrieval technology integrating the visual, movement and audio features can set the model for the spectacular Rally scenes in tennis match video. We can retrieve the same which is in line with most people's thoughts through the spectacular degree model of the Replayed Rally scenes and the spectacular Rally scenes.

References

1. LY. Xing, QX. Ye, WG. Zhang. A scheme for racquet sports video analysis with the combination of audio-visual information. In: Proceedings of International Conference on Visual Communications and Image Processing, Vol. 5960, 2005.
2. Lin Tong, Zhang Hong Jiang, Feng Jufu, etc. Analysis on Contents of Shots and Application of Contents of Shots in Video Retrieval. Journal of Software, 2002, 13(8) , pp. 1577-1585. (in Chinese)
3. Chong-Wah Ngo, Yu-Fei Ma, Hong-Jiang Zhang. Video Summarization and Scene Detection by Graph Modeling, IEEE Transactions on Circuits and Systems for Video Technology, 2005, Vol. 15, No. 2, pp. 296-305.
4. Ma Y F, Zhang H J. A model of motion attention for video In Proceeding of IEEE International Conference on Image Processing, Rochester, New York Sep, 2002, pp. I29-132.
5. Wu Fei, Zhuang Yueting, Pan Yunhe, etc. Identification and Retrieval of Audio Samples Based on Incremental Learning of SVM. Journal of Computer Research and Development, 2003, 40 (7) , pp. 950-955. (in Chinese)
6. Zhuang Yueting, Liu Junwei, Wu Fei, Pan Yunhe, Zhang Yin. Automatic Caption Location and Extraction Based on Support Vector Machine. Journal of Tsinghua University, 2002,14(8) , pp. 750-753. (in Chinese)