Research on Online Sports Metadata Extraction System based on Video Processing Technology

Haixin Yao[*1] and Jinmei Shao[2]

[1]School of Physical Culture, North China University Of Science And Technology, Tangshan 063009, China
[2]School of Physical Culture, North China University Of Science And Technology, Tangshan 063009, China
yaomaomao_217@sohu.com

Abstract

Sports video metadata extraction system based on the content of basic goal use an automated or semi-automated interactive means to obtain video data as complete features and attributes for efficient retrieval mechanism. For fast access to video information needed, sports video ornamental create conditions. Firstly, video-based layered metadata description model, we discuss the structure of the video processing technology, and an increase in the time domain and airspace video object motion information on this basis. Low-level visual features for video and high-level semantic features presents a particular field of video information for video implicit hierarchical division method. Video automated visual feature extraction, semantic feature places marked attracted achieve human-computer interaction. Focus on the sports information descriptors and visual content descriptors, descriptor structure video. Video data based on hierarchical structure model and video features standard video content description model.

Keywords: Applied research; video metadata; online video; detecting video segmentation; feature space

1. Introduction

With the accumulation of television video programs, the online digital video increased, as well as video on demand. Distance learning and other large media popularity of multimedia access technology has become more abundant. Multimedia information with an intuitive and friendly, wonderful and rich features, but also has the characteristics of data processing is not conducive to large, unstructured, meaning more content and so on [1]. Wherein the video information is multimedia information in the largest proportion, up to the amount of information, the most widely used application prospect of complex parts [2]. By analyzing the structure and content features sports videos automatically recognize extract highlight scenes interested viewers can save time viewers to watch sports [3]. Users can quickly and efficiently find the desired video from a lot of content, directly and accurately locate and browse, time-saving, convenient and quick. In addition, if the game sports great shots assigned a lower compression ratio and higher transmission rate, other fragments to give a higher compression ratio and lower transmission rate, so that you can give users a sense of adaptively content of interest, providing personalized video transmission, and increase the transmission efficiency of video coding [4, 5].

Microsoft's RCE scheme, metadata editing is also an important feature [6, 7]. In order to allow the end user when playing video, it is possible to simultaneously obtain more information about the video, while making video clips, also you need to edit the metadata [8]. Video metadata development so far, we need to focus on more than just the contents of the properties and description of these two types of metadata, metadata and asked hotspot class metadata to provide more details of the video information, concerned about

the user experience today, metadata position in the video clip system without saying [9]. In this paper, "Application of online sports video editing system metadata," a title, aims to combine the existing research information on sports video metadata for online video editing system proposed its own set of metadata solutions. Thesis, study video metadata format has theoretical depth, metadata editing application value in the realization of online video editing system, for sporting events such video content on metadata handling a certain innovation.

The main work of this paper is to describe the model sports video information and video metadata extraction technology research and design and implementation of sports video metadata extraction system. Reviewing and summarizing the results have been based on the object description field in sports video presents a new multi-level sports video content metadata description model based not only on the structure of the metadata standard definition video is carried out and described, but also for video object metadata also established a new description model.

2. Related Theory and Method

2.1 Sports ML

Sports is using a set of IPTC Metadata XML standard developed primarily for switching and transmission of all kinds of sports Records and related news stories [10, 11]. Sports ML features to provide the ability to identify and describe a large number of sports information, the focus of which stressed that [12]: (1) score (Scores), (2) schedule (Schedules), (3) Places (Standings), (4) a record Statistics (Statistics) and (5) News (News) and other five within the Core DTD Sports ML develop a large number of sports news stories describe the characteristics of the majority of sports news stories can be written using Core DTD, in addition, in order to allow news reporters We can be more in-depth coverage of sports news [13]. Sports element is divided into several parts, briefly described in Table 1.

Table 1. Sports ML Element Description Table

Brief name	Explain
Sports-Metadata	Mainly describes the Sports ML file number, theme, characteristics, classification, summary, etc.
Sports-Event	Describes the results of the sports competitions, including the referee, the type of sports activities, important events, the type of medal and the results of the relevant competition.
Tournament	Describe all kinds of details of the round robin tournament, including the background, the local information, the competition structure, etc.
Schedule	Describe the race schedule
Standing	Description of the team or individual players ranking table
Statistic	Statistics and comparison of the records of the team or individual players
Article	Describe the sports activities refer to the relevant news reports, in order to recommend NITF format.
Player	Description of the basic information of individual players, including name, height, weight, nationality, etc.
Team	The basic data description of the team, including the identification of whether home or away.

Individual sports have their own unique set described manner, Sports ML as a data exchange platform, it must take into account the different record sports [14]. IPTC develop SportsML main goal is to pass the history of sports, and news reports at the same time be combined, as in the exchange of information is an important normative standards, therefore SportsML not only can be used independently to describe sports record, and more capable of NITF NewsML integration to deliver during the news or sports information can flow more exchange on different platforms, different languages, so as a descriptive markup language sports, SpoasML addition to describing the content of sports, including a description of the player or basic information matches the team, schedule, rankings, records statistics, but also can be described in a single sports event, then the event described in reference to NITF or details related news content, and finally to carry out NewsML news package, in order to overall structural and integrity.

2.2 Sports Video Metadata Model

In order to meet the needs of online sports video editing system, refer to various existing video metadata standard features, they complement each other. Selected the practical value of the portion of the integrated system is applied to the paper's sports video metadata model [15]. This video-based content metadata model needs to be well supported video retrieval, storage and subsequent rendering, so compatible with internationally accepted standards, can describe specific sports, support temporal metadata descriptions.
Global metadata is a global sports video file description, once the video file is edited, the example in this file to a video clip with another video clip files in a merger epigenetic became the new video file, then the global metadata of the original video file is not inherited the new video file; and temporal metadata itself has a characteristic passage, as would normally be the basis of a video clip, and should be retained in the new generation of video file. After the integration of the various drawing on existing video and metadata standards, the formation of a sports video metadata model in this paper, the structure was shown in Figure 1.
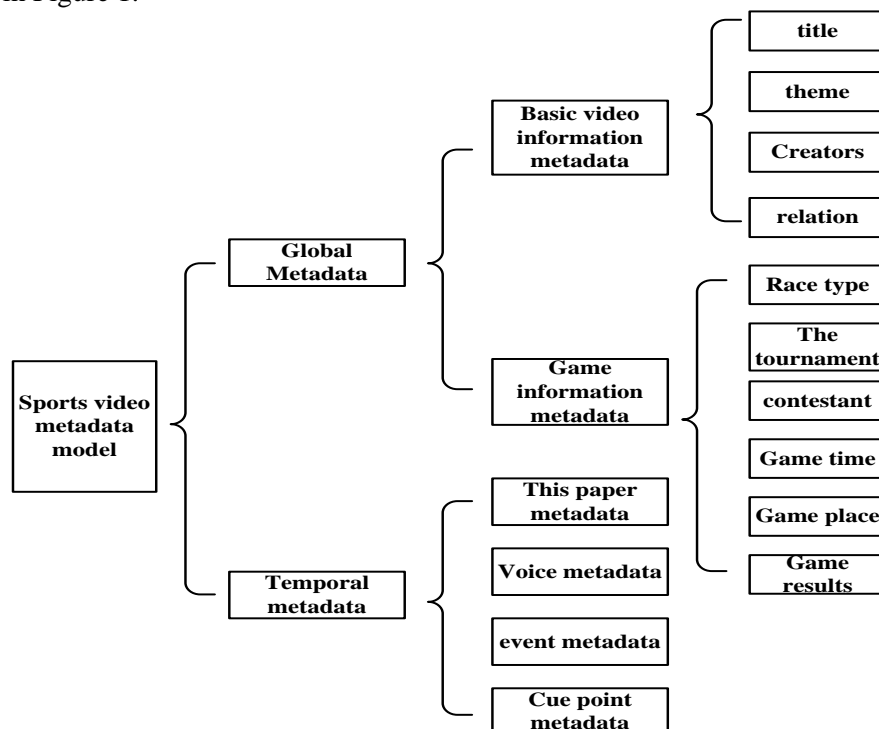

Figure 1. Sports Video Metadata Model Structure Diagram

Figure 1, in the sports video metadata model, the metadata into the global metadata and temporal metadata in two parts. Among them, the global metadata can be divided into the basic information metadata and video game information metadata. Basic video information metadata mainly with reference to DC coding standards and MEPG-7 standard, they keep those labels on video basic information described in the figure lists only a few major labels, partly to ensure that the existing standards and formats Inheritance and compatible. Global competition metadata is a reference to the Sports ML standards for these special sports video content in more detail - in a field of sports events (sporting events or fragment) is described. The temporal metadata different from the global metadata, they have the timing, in each class metadata record, must contain the start time, end time (or duration) and other labels to describe its timeliness.

2.3 Video Feature Selection

Color histogram can be based on different color space and coordinate system. Some typical common coordinate system such as: RGB color coordinate system, YUV color coordinate system, HSI color coordinate system and the like. HSI color coordinate system due to the direct correspondence with hue, saturation, and brightness of the three elements of human vision, the use of this coordinate system can greatly simplify the workload of image analysis and processing. Studies have shown that, H does not have the space with brightness, contrast, and advantages of the change occurred, it is useful for the color extraction.

Like imaging systems use RGB color coordinates, color images can also use this color coordinate system f (x, y). Contains three color components conclude a red component R (x, y), green component G (x, y) and blue component B (x, Y). Luminance values obtained through the color component weighted formula (1) as follows:

$$f(x, y) = a_1 R(x, y) + a_2 G(x, y) + a_3 B(x, y) \tag{1}$$

In which $a_i(i = 1, 2, 3)$ represents the sensitivity of the eye color.

Color television systems generally use YUV color, the color of which it will break down into luminance and color difference signals: Y represents luminance information of the color; the U and V color difference signal. YUV coordinate conversion formula and RGB coordinate system such as formula (2) below:

$$[Y \quad U \quad V] = [R \quad G \quad B] \begin{bmatrix} 0.223 & -0.125 & 0.532 \\ 0.468 & -0.523 & 0.581 \\ 0.825 & -0.915 & 0.193 \end{bmatrix} \tag{2}$$

The coordinate system is divided into color hue, saturation and brightness. HVS three elements correspond to. Hue H is reflected by the object or transmitted through an object's color, hue is identified by the color name from, such as red, orange or green; Saturation, sometimes called chroma, it refers to the color intensity or purity. Saturation, hue represents the ratio of the share of the gray component as a percentage from 0% (gray) to 100% (fully saturated) to measure. Brightness is the relative lightness or darkness of the color, usually a percentage from 0% (black) to 100% (white) to measure. HIS to RGB conversion formula such as formula (3) (4) (5) (6):

$$I = (R + G + B) / 3 \tag{3}$$

$$S = 1 - \frac{1}{I} \min(R, G, B) \tag{4}$$

$$H = \begin{cases} \theta & G \geq B \\ 2\pi - \theta & G \end{cases} \tag{5}$$

$$\theta = \arccos\left\{\frac{[(R-G)+(R+G)]/2}{(R-G)(R-G)+(R-B)(R-G)}\right\}$$

(6)

Texture surface structure of the organization to reflect the arrangement of objects, which is not dependent on the color or brightness of the visual features, color feature comparison of its more represent. So the texture features in image retrieval, pattern recognition and computer vision developed rapidly. Autoregressive Texture Model SAR is based MRF model is successfully applied. In the SAR by S represents a pixel intensity values g (s) can be expressed as a linear superposition of its adjacent pixel intensity values with the noise term ε (s) and the specific formula is as follows (7) shown below:

$$g(s) = \mu + \sum_{r \in D}^{a} \theta(r) g(s+r) + g(s)$$

(7)

For large and complex XML documents it is ideal, because the data is structured. This not only allows the user to specify a definition of vocabulary elements in the document, but you can also specify the relationship between elements. Basically speaking, XML is a document format. It is a series of rules on XML document what it looks like, the degree of compliance with the XML standard has two levels: the first level is the structural integrity, the second level is correct.

3. Video and Metadata Management

3.1 Storing Metadata Information

For half way to preserve the merger discussions store metadata information, and now more mainstream two storage methods are: the use of relational database storage and use of XML file storage. Early relational database using similar systems use a relational database to store metadata, they use the relationship between data tables to describe the relationship between the metadata and video files, and records in the table of contents shows the contents of metadata information. You can use the following conceptual data model is used to describe the relationship between the various types of video data is detailed in Figure 2. The figure lists only the most important relationship between the three entities.

Video Table			
ID	<pi>	Interger	<M>
Name		Variable characters(200)	
Path		Variable characters(200)	<M>
Title		Variable characters(500)	
Theme		Variable characters(500)	
Creator		Variable characters(200)	
Entry Time		Timestamp	<M>
Key Identi	<pi>		

MetaData Table			
ID	<pi>	Interger	<M>
Type		Variable characters(20)	
Start Time		Timestamp	<M>
End Time		Timestamp	
Duration		Number(20,2)	
Description		Text(1000)	
Key Identi	<pi>		

Match Table			
ID	<pi>	Interger	<M>
Type ID		Short interger	
Name		Variable characters(200)	
Participant		Variable characters(500)	
Time		Date&Time	
Location		Variable characters(200)	
Result		Variable characters(200)	
Key Identi	<pi>		

Figure 2. Conceptual Data Model Diagram

Video file name table used to describe video files, paths, title, subject, author, when asked warehousing and other basic information, as well as video files associated with the game, the game can have multiple video files. Association table used to describe the correspondence between video files and metadata information, a video file may correspond to a plurality of metadata information recording.

Online sports video editing system, the storage of metadata is mainly used in this way, the associated XML document in a file system, use the file system to manage them. In a sense, XML file itself can be seen as a database or a table in which, because of its self-description, unlimited nested tree structure and other characteristics. Use XML files store information using a database to store information and there is a relatively large difference is: can belong to a variety of video file metadata information into focus after a single XML document classification, unified format through the label, the dispersion of information integrated together.

3.2 Search Results for Many Types of Information Gathering Process

Using a single source of information search out the results are often not comprehensive, because each type of information has its limitations: Subtitle and video playback content information may not match (for example, the bottom of the screen scrolling marquee), and not every time the key event information is relatively full, but may not have a user searches for the information in the description; every moment has a corresponding caption information; audio information, events guide topics mentioned at some point likely to be completely okay with the current situation of the game word, which can not be searched.
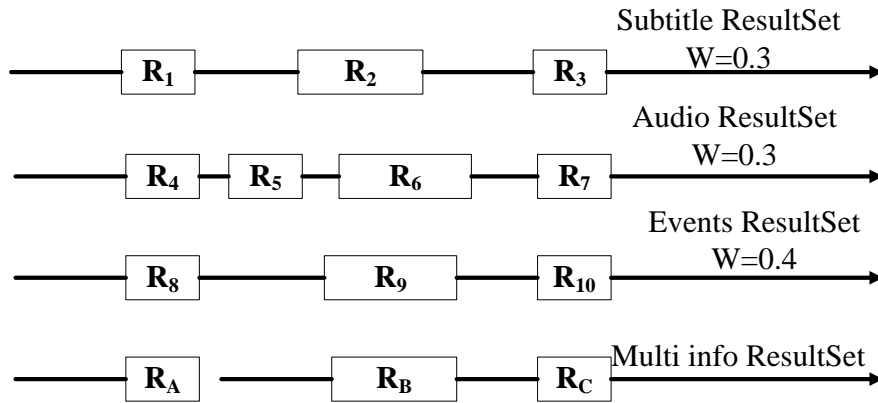
Figure 3. Search Results for Many Types of Information Gathered Schematic

As shown in Figure 3, the search results were aggregated information from different sources, you may get the result set is not the same. If you compare them in the time dimension, easy to see, at R3 subtitle information it is likely to be unrelated to the video marquee, R, instructors at the event tells a topic and then race without too much association. In these three types of information gathered result set again, the need for three types of information separately set the weight (w), as selected or not when conflicts determination conditions, which would be described as follows: Choose the heaviest weight class that information as a reference, and the results dimension and ask the other results set information obtained by comparing the reference correction when the result set. According to the characteristics of different sources of information, there is at the repair 'F' the following principles: 1) Select the priority at the start of the event information result Q; 2) select the priority of audio information results at the end of Q; 3) when the weight of an outcome heavier combined weight is less than the threshold value when Φw (Φw default is 0.6), concentrated to remove this from the results found.

4. Experiment and Results

4.1 Shot Detection

Shot Detection is also known as a scene change detection, scene change into two categories: shear and gradient. Gradient is divided into rules gradient (fade, dissolve, pan, etc.) and irregular gradient. Generally, between frames within a shot with little difference, different lenses inter quite different. The main difference in color characterization object has become larger, far away from the edge of the old and new, object shape changes and movement of small continuity. The main algorithm: a method based on the comparison of the frame, the method based on edge detection. Showing similarity with two frame distance D, D smaller the similarity of two higher, the lower the likelihood of a scene change, and vice versa, when D is above the threshold T when the camera switches can be considered to exist. Inter-pixel distance may be a distance D, the specific formula is as follows:

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

(8)

Wherein, Xi represents a video frame i, d (Ix, Iy) and the inter-Ii 'distance, Ii (x, y) is the i-frame pixel values (x, y) location, M and N of the frame width and height. This approach corresponds to compare before and after two changes between pixels, if the change exceeds a threshold, then that lens switching. Frame distance D may also use pixel brightness and color statistics, specific formula is as follows:

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

(9)

Service layer three services listed in this system is mainly involved in these services by the server to a specific system implementation. Content Management, Indexing Service is responsible for management and maintenance of video files associated metadata file each logical path (URL) and a physical path between video files and associated metadata files, and supports video-based content inspection. Project Management Service is responsible for managing the project when the user uses the system created. For exports of services, the project file reception system-generated video EDL will be extracted, a plurality of video clips in the service was really encoded into new video file, the corresponding meta-data and information will be saved associate. Figure 4 is a one-minute video shot detection. The ordinate indicates the inter-frame distance.
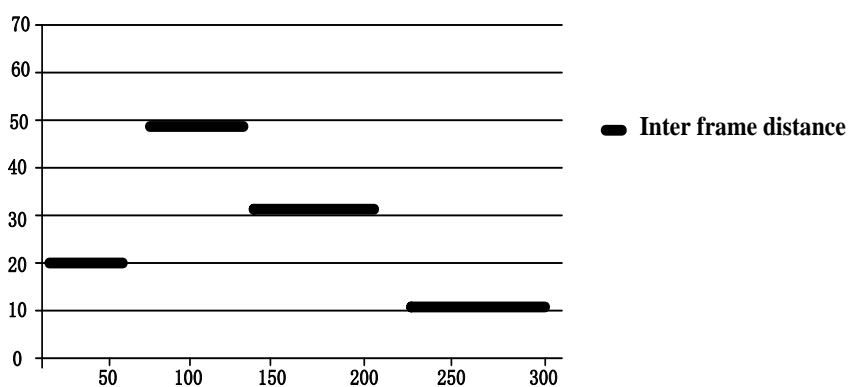
Shot detection

Figure 4. Comparison Shot Detection Frame

Max (din, dout) greater than a certain value reading is considered that a lens switching. Figure 5 shows a one-minute video edge detection ordinate min (din, dout).
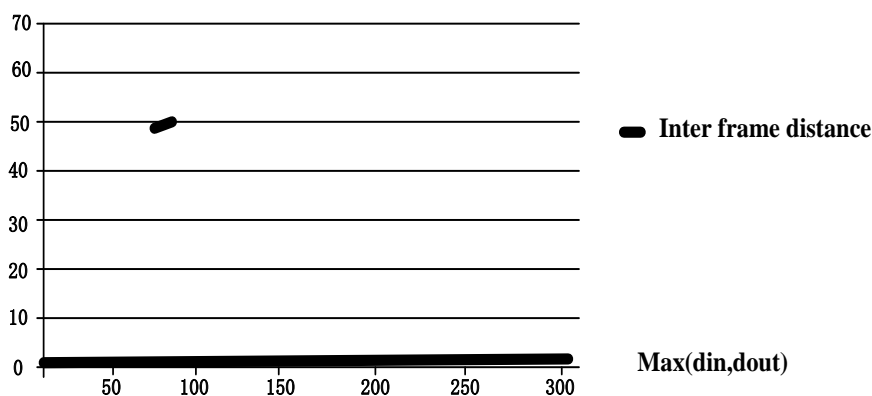
Shot detection

Figure 5. One-minute Video Edge Detection Ordinate

Keyframes can be used to describe a lens that can reflect the main content of a representative of a shot, select the keyframe on the one hand to be able to reflect the major events of the lens, which should be described as accurately as possible completely; on the other hand, in order to facilitate management should have a small amount of data, and the calculation should not be too complicated. Many key frame selection method, a common frame averaging and histogram averaging method. Frame averaging is to take the average

of all the pixel values in a frame from a position on the lens, and the lens frame of pixel values in the position closest to the average value as a key frame; histogram averaging sucked all the shots histogram statistics frame averaging, and then select the average histogram closest frame as a key frame. The advantage of these methods is relatively simple calculation, the selected frame has an average representative of significance; drawback is to select a keyframe from one shot, you can not describe a plurality of motion of the lens. The number of dynamically allocating keyframes to change to avoid a few too many shots selected key frames, and for sports more shot with a key frame and can not adequately describe.

4.2 Full-text Search of Sports Video Metadata

For it to achieve an effective and efficient sports video metadata framework for full-text search, you need to run properly designed system architecture, simultaneously using an efficient search engine process, and get the search results to further improve the hit rate of search processing. As used herein, the sports video metadata to achieve full-text search system consists of three parts: the collection and pretreatment (1) of the original data information; (2) Solr full-text search service; (3) pre-processing and presenting search results. System structure was shown in Figure 6.

Resource load module call content management services layer, indexing service, is responsible for obtaining from the content manager to be edited video files and their metadata. Output module calls the service layer exports of services, responsible for the results generated video clips. Video clip module as the main modules interact with the user, contains four modules: Library sub-module, sub-module content viewer, preview the results of the sub-modules and sub-modules when asked axis, back in these modules, users can easily video clips and metadata editing work.

Sports video content-based metadata extraction system goal is to use the method described above, video content analysis and automatic man-machine interactive annotation, extract a clear hierarchy, semantic information-rich sports video metadata. Metadata is stored in XML format, the video format and the physical storage unrelated and independent existence, consistent temporal and spatial domain by a clear description and video for each level of time and space to reach.
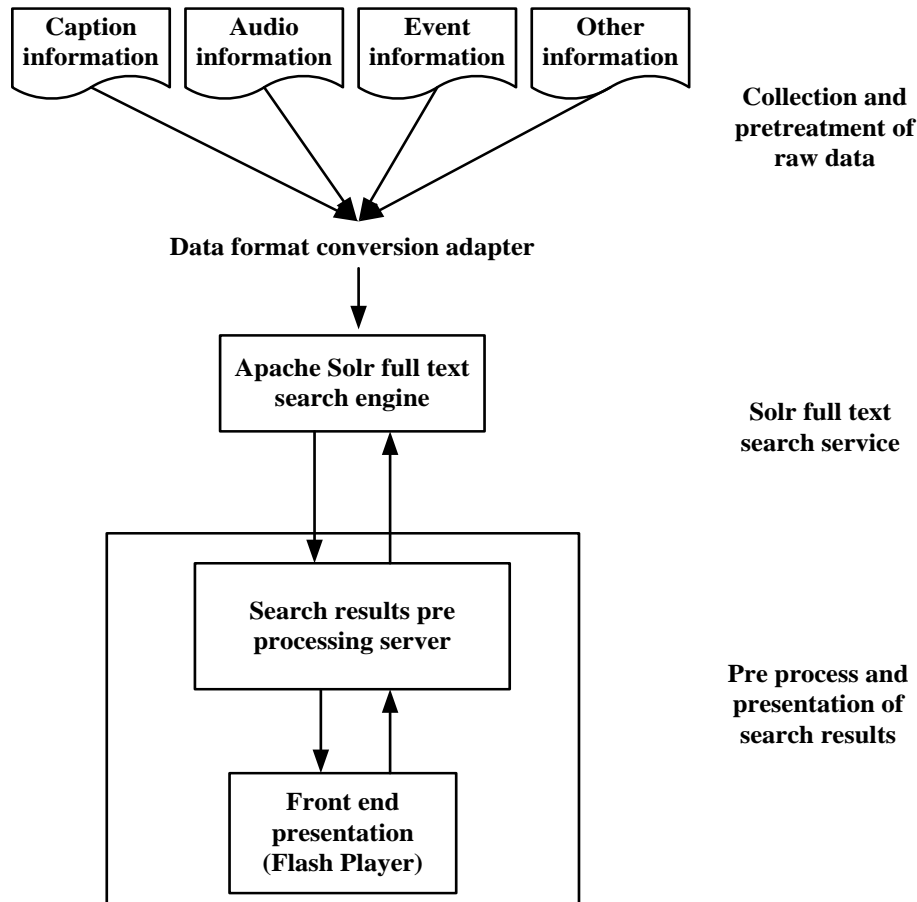
```
┌──────────┐ ┌──────────┐ ┌──────────┐ ┌──────────┐
│ Caption  │ │  Audio   │ │  Event   │ │  Other   │
│information│ │information│ │information│ │information│
└──────────┘ └──────────┘ └──────────┘ └──────────┘
```

Collection and pretreatment of raw data

Data format conversion adapter

```
┌──────────────────────┐
│ Apache Solr full text │
│     search engine     │
└──────────────────────┘
```

Solr full text search service

```
┌──────────────────────┐
│ Search results pre    │
│  processing server    │
└──────────────────────┘

┌──────────────────────┐
│    Front end          │
│   presentation        │
│   (Flash Player)      │
└──────────────────────┘
```

Pre process and presentation of search results

Figure 6. Sports Video Metadata Structure of Full-text Search System

5. Conclusions

This article focuses on sports video metadata applications in online sports video editing systems, and research results from existing systems starting detailed design format video metadata, storage and realization of metadata framework for full-text search, online sports editing video editing system metadata, as well as client-side rendering. Sports video metadata extraction system based on the content of the basic goal is to use an automated or semi-automated interactive means to obtain video data as complete as possible features and attributes for efficient retrieval mechanism for fast access to video information needed to increase Sports video ornamental create conditions. Firstly, video-based layered metadata description model, we discuss the structure of the video processing technology, and an increase in the time domain and airspace video object motion information on this basis. Low-level visual features for video and high-level semantic features presents a particular field of video information for video implicit hierarchical division method. Video automated visual feature extraction, semantic feature places marked attracted achieve human-computer interaction.

Acknowledgments

References

[1] R. M. Perianu, M. M. Perianu, P. Havinga, S. Taylor, R. Begg, M. Palaniswami and D. Rouffet, "A performance analysis of a wireless body-area network monitoring system for professional cycling", Personal and Ubiquitous Computing, vol. 17, no. 1, (2013), pp. 197-209.

[2] P. Halvorsen, S. Sægrov and A. Mortensen, "Bagadus: an integrated system for arena sports analytics: a soccer case study", Proceedings of the 4th ACM Multimedia Systems Conference. ACM, (2013), pp.48-59.

[3] M. Justin, T. Grossman and G. Fitzmaurice, "Video lens: rapid playback and exploration of large video collections and associated metadata", Proceedings of the 27th annual ACM symposium on User interface software and technology. ACM, (2014).

[4] T. Saba and A. Altameem, "Analysis of vision based systems to detect real time goal events in soccer videos", Applied Artificial Intelligence, vol. 27, no. 7, (2013), pp. 656-667.

[5] R. Barthel, S. Ainsworth and M. Sharples, "Collaborative knowledge building with shared video representations", International Journal of Human-Computer Studies, vol. 71, no. 1, (2013), pp. 59-75.

[6] C. Dowd, "The new order of news and social media enterprises: visualizations, linked data, and new methods and practices in journalism", Communication Research and Practice, vol. 2, no. 1, (2016), pp. 97-110.

[7] D. O'Callaghan, D. Greene and M. Conway, "Down the (White) Rabbit Hole the Extreme Right and Online Recommender Systems", Social Science Computer Review, vol. 33, no. 4, (2015), pp. 459-478.

[8] S. P. Algur, P. Bhat and S. Jain, "Metadata Construction Model for Web Videos: A Domain Specific Approach", International Journal of Engineering and Computer Science, (2014).

[9] C. Gkonela and K. Chorianopoulos, "VideoSkip: event detection in social web videos with an implicit user heuristic", Multimedia Tools and Applications, vol. 69, no. 2, (2014), pp.383-396.

[10] A. Rehman and T. Saba, "Features extraction for soccer video semantic analysis: current achievements and remaining issues", Artificial Intelligence Review, vol. 41, no. 3, (2014), pp.451-461.

[11] C. Tzelepis, Z. Ma and V. Mezaris, "Event-based media processing and analysis: A survey of the literature", Image and Vision Computing, (2016).

[12] A. Bermingham, N. Caprani, R. Collins, "Recommending Video Content for Use in Group-Based Reminiscence Therapy", Health Monitoring and Personalized Feedback using Multimedia Data. Springer International Publishing, (2015), pp. 215-244.

[13] A. Merlino, J. S. Berg and B. Dahi, "Rapid triage of diverse media devices: Boston marathon bombings: An operational case study of time critical response", Technologies for Homeland Security (HST), 2013 IEEE International Conference on. IEEE, (2013), pp.207-211.

[14] K. Chorianopoulos, D. A. Shamma and L. Kennedy, "Social Video Retrieval: Research Methods in Controlling, Sharing, and Editing of Web Video", Social Media Retrieval. Springer London, (2013), pp. 3-22.

[15] A. Karahasanović and J. Heim, "Understanding the behaviour of online TV users", Personal and Ubiquitous Computing, vol. 19, no. 5-6, (2015), pp. 839-852.

Authors

Haixin Yao. Female, born in February 1979, Tangshan, China. She received her Master Degree from School of Physical Culture, TianJin University, TianJin, China in 2008. Currently, She is a lecturer working in School of Physical Culture at North China University of Science and Technology. She has published more than 50 academic papers, her current research interests include Pattern Recognition Algorithm and Physical Education.

Jinmei Shao. Female, born in 1970, yancheng, jiangsu province. She received her Master Degree from School of Physical Culture, Shanghai University, Shanghai, China in 2003. Currently, She is an associate professor working in School of Physical Culture at North China University of Science and Technology. She has published more than 60 academic papers, her current research interests include Pattern Recognition Algorithm and Sports Training.