Content-Based News Video Mining*

Junqing Yu[1], Yunfeng He[1], and Shijun Li[2]

[1] Computer College of Science & Technology,
Huazhong University of Science & Technology , Wuhan, 430074, China
yjqinghust@126.com, heyunfeng@tom.com
[2] School of Computer, Wuhan University, Wuhan, 430072, China
shjli@public.wh.hb.cn

Abstract. It is a challenging issue to analyze video content for video mining due to the difficulty in video representation. A hierarchical model of video representation is proposed with a schema for content-based analysis of news video in this paper. The research problem targeted in this paper is to mine a massive video database to retrieve specific clip based on content defined by users. This is frequently encountered in entertainment and video editing. A novel solution to this problem is developed in this paper, in which the consecutive news video is segmented into shots, scenes and news items using multimodal features based on the hierarchical model. To summarize the content of video, a video abstract is developed. The experimental evaluation demonstrates the effectiveness of the approaches discussed in this paper.

1 Introduction

With the development of multimedia and web technology, the multimedia data, including image, audio and video, have been produced massively. Digital video rapidly becomes an important source for information, education and entertainment. It is needed urgently the advanced technologies for organizing, analyzing, representing, indexing, filtering, retrieving and mining the vast amount of videos to retrieve specific information based on video content effectively, and to facilitate new and better ways of entertainment and multimedia applications. Content-base video analyzing and retrieval are important technologies, which have been an international research focus in recent ten years. As challenging problem, content-based video mining is also emphasized by lots of researchers. Although numerous papers have been published on data mining [1, 2], few of them deal with video mining directly [3, 4]. Low features, such as color, texture, audio and motion, can be used to segment video sequence into shots and to extract caption or other region of interest for video data management and retrieval. However, the mining of video data based on its content is still in its infancy. Due to the inherent complexity of video data, existing data mining algorithms and techniques can not be used directly in video data. The new mining techniques or modified ones should be designed to facilitate the video data mining process.

* This paper is financially supported by Natural Science Foundation of Hubei Province (2004AA101C94).

X. Li, S. Wang, and Z.Y. Dong (Eds.): ADMA 2004, LNAI 3584, pp. 431–438, 2005.

Generally speaking, there are two kinds of videos in our daily life [5]: videos with some content structures and videos without any content structure. The former are videos such as movies and news where scenarios are used to convey video content. The latter is like surveillance videos, they have no scene change, therefore no content structure can be found among them. Just because of these, we can not process video data using a unified approach like dealing with text data. Specific processing schema should be designed for the different kinds of video data and many efforts had been made.

In today's society, the amount of news information generated is growing exponentially. Moreover, the data is made available in more than one dimension across different media such as video, audio, and text. This mass of news information poses serious technological challenges in terms of how news data can be integrated, processed, organized, and indexed in a semantically meaningful manner to facilitate effective retrieval. Because of its usefulness and importance, there have been many research efforts in news video analysis. R. Mohan [6] proposes to segment TV news by synchronizing images with the associated close-captions or teletext (the European version of close-captions). L. Chen [7] presents multi-criteria video segmentation based on image and sound analysis. Zhang et al [8] base their work on the anchorperson position in order to split the news into independent subjects. This model fits a type of news where the anchorperson and camera position do not change much. The Informedia's work is very impressive, in which speech recognition and image analysis were combined to extract content information and to build indexes and abstracts of news video [9, 10]. Lately, many researchers adopted the idea that image, audio and speech analysis are integrated in video content analysis [11, 12].

The rest of the paper is organized as following. Section 2 proposes a hierarchical video organization schema. A news video sequence is segmented into shots and news items based on audiovisual features in Section 3. A novel mining tool, news video abstract based on key frames, is introduced in Section 4. Section 5 concludes this paper.

2 Hierarchical Video Organizing Schema

For the video with content structure, video data usually bear hierarchy in both content and structure. Accordingly, a hierarchical video organizing schema is introduced and an independent object identifier can be assigned to every video object.

2.1 Hierarchical Video Organizing Model

The original video, with content structure, can be organized in five levels: video event, episode, scene, shot and frame image. All but the shortest video are made up of a number of distinct scenes, each of which can be further broken into individual shot depicting a single view, conversation or action. A shot designates a continuous sequence of frames, which are bottom level of the model and are corresponding to the temporal image sequence of the original video. Using high level semantics, some scenes (neighboring or not) can be combined into episode. Episode makes up the semantic unit of video and depicts a story or an action. In the same episode, the content of scenes is relevant, but they can be separated in temporal order.

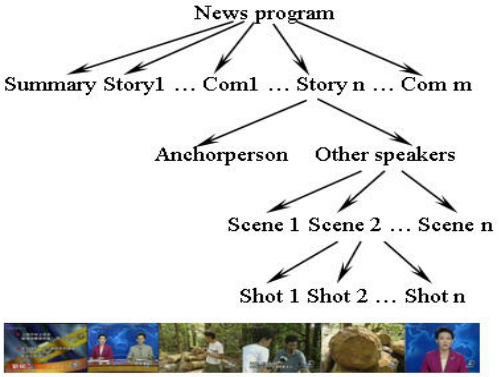

Fig. 1. Hierarchical organization of news video

2.2 News Video Organizing Schema

We can use hierarchical model to organize news video. Fig.1 shows the hierarchical organization of news video. A typical national news program (e.g., CCTV news) consists of news and commercials. News is made up of several headline stories, each of which is usually introduced and summarized by the anchor prior to and following the detailed report by correspondents, quotes, and interviews from newsmaker. Commercials are usually found between different news stories. Based on this hierarchy of news video, we propose a prototype of news video database system, with which continuous news video can be automatically analyzed based on content by utilizing different cues of audiovisual information.

3 Content-Based Analysis of News Video

To retrieve or mine news video data, one of the most important tasks is to the transform the original video sequence into a hierarchical dataset according to the model depicted in Fig. 1. To facilitate this goal, we adopt some video processing techniques to analyze the news video. Users have different needs when mining news data. For instance, some users may want to directly retrieve a news story; some may like to listen to the news summary of the day in order to decide which story sounds interesting before choosing what to watch further. In order to satisfy different requirements, a segmentation mechanism is needed that partitioning news video data in different ways so that direct indices to the events at different levels of abstraction can be automatically established.

3.1 Feature Extraction

Digital video is a dynamic sequence, which contains image and audio signal simultaneously. Compared with the text, static image and audio, the video's content is much more complicated and abundant. However, restricted by the present computer tech-

nology and artificial intelligence, high-level semantics cannot be directly extracted from video and low-level features should be extracted firstly. Three sources of information can be used for video processing. They are color, audio and motion. Most of the existing video segmentation algorithms are based on visual information, such as color histogram, edge feature, etc. Combined audio and visual approaches have been considered only recent years. Here, we propose a new feature extracting approach, which is based on Microsoft DirectShow SDK system. Special audio-visual feature extractors, which are custom filters in DirectShow, are designed to extract features in real time. Extracted features are listed in the following [12, 13]:

Audio features: NSR (non-silence-ratio), ZCR (zero crossing rate), STE (short-time energy), VDR (volume dynamic range), VU (volume undulation), FC (frequency centroid) and BW (bandwidth);

Color features: DC (dominant color), DCStd (standard deviation of dominant color), PDC (percentage of dominant color), Color Histogram, DCH (mean of difference between the color histograms);

Motion features: DMX (X component of dominant motion), DMY (Y component of dominant motion), PDM (percentage of dominant motion), ME (mean motion energy).

3.2 Anchorperson Frame Detection

The main content of news video is a series of news story, so exactly detecting the boundary between news stories is very important for the content-based mining news data. It is not difficult to find that the anchorperson frame is usually the beginning or ending frame of individual news story. Just because of this, automatically detecting the anchorperson frame can help recover the news stories. Most of the existing approaches to this problem are either based on face detection or based on speaker identification, so their computational complexity is very great. Here, a completely new and simplified method was proposed. The detailed detection process can be explained as follow:

Step 1: Establishing the DC template of anchorperson frame. For most of news video, the first anchorperson frame usually appears after the theme music in a relatively rigid time interval, the template frame can be automatically chose through the detection of theme music, which has been discussed in [14]. Suppose M frames are chose to be template frame, the DC template can be computed by equations 1 and 2.

$$DCStd_T = \frac{1}{M}\sum_{i=1}^{M} DCStd_T(i) \tag{1}$$

$$PDC_T = \frac{1}{M}\sum_{i=1}^{M} PDC_T(i) \tag{2}$$

Where $DCTStd_T(i)$ and $PDC_T(i)$ indicate the standard deviation and percentage of dominant color in i^{th} template frame. $DCTStd_T$ and PDC_T stand for the template features.

Step 2: Computing the DCStd and PDC features of every frame image in the news sequence.

Step 3: Template matching. We compute the difference, *D(i)* ,between the template feature and responding feature in i^{th} frame by equation 3. If *D(i)* is less than the predefined threshold, which can be adjusted adaptively, the frame can be identified to be anchorperson frame and it can be marked automatically.

$$D(i) = \sqrt{C_1(DCStd_i - DCStd_T)^2 + C_2(PDC_i - PDC_T)^2} \quad (1 \le i \le N) \tag{3}$$

Where C_1 and C_2 indicate the weight of DCStd and PDC feature, N refers to the reframe number in the news sequence. If in RGB color model, we can compute the DCStd and PDC in red, green and blue separately.

3.3 News Story Segmentation

In the news video, a period of news program usually consists of several news stories, and the news story is made up of some scenes or shots. Therefore, the approach to parsing shot and scene can be used here. To detect news story boundaries, robust anchorperson frame detection is needed, because the frame with anchorperson is often appeared at the beginning of news story. The silence between stories is also important cue. Just based on these considerations, approach based audio-visual information is proposed [14, 15]. The entire segmenting process can be divided into two steps, one is to search candidate boundary points, and the other is to verify the candidate boundary points.

Step 1: Searching candidate boundary points. Continuous news video sequence consists of two types of clip, one is anchorperson frame chip, and the other is non-anchorperson frame chip. Using the approach discussed in Section 3.2, we can find the candidate boundary points conveniently. Fig.2 gives illustration of 30 minutes news video, in which SC1, SC2, … , and SC12 stand for candidate points.. Among the candidate points, some are not exactly the boundary points of news story, and we call them false points. They maybe belong to the same news story. On the other hand, some true boundary point might be ignored if no anchorperson frame exists in some news story, or one story ends by anchorperson frame and the next neighboring story begins with anchorperson frame. Therefore, we have to verify the candidate point in the next step.

Fig. 2. Candidate boundary points of news stories

Step 2: Verifying the candidate boundary points. The objective of this step is to delete false points and supplement ignored ones. Here, the silence clip between the news stories was utilized. The short-time zero (STZ) rate is used to detect the silence chip, Fig. 3 gives a silence detecting result.

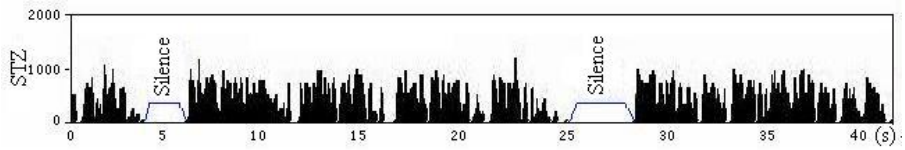

Fig. 3. Detection result of silence chip between news stories

For convenience, we can use a dualistic group to express the detected silence chip like equation 4.

$$SG(i) =< s_i, e_i >, i, s_i, e_i = 1,2,\cdots, and\ s_i \le e_i \tag{4}$$

Where, s_i and e_i indicate the starting and ending frame number of i^{th} silence chip. Two theorems are used to finish the verifying process as followed.

Theorem 1: If no silence chip exists in one candidate boundary point, this is false point and should be deleted.

Theorem 2: If one silence chip, $SG(j)=<s_j, e_j>$, exists in one anchorperson frame chip, there is a boundary point at this silence chip and it should be supplemented. Suppose this supplemented point to be $SB(j)$, it can be computed by equation 5.

$$SB(j) = (s_j + e_j)/2 \tag{5}$$

Table 1. The experimental results of news story segmentation

News video	Actual	Detected	False	Missed
News reports	44	45	1	0
Night news	38	38	1	1
Total	82	83	2	1

Through the above two steps, all the true boundary points have been identified, and using them we can segment a continuous news video into a series of news stories. Table 1 gives the experimental results of news story segmentation for CCTV news reports and night news. The results have revealed the effectiveness of segmentation algorithm.

4 News Video Abstract Based on Key Frame

The summarization of video content provides an effective way to speed up video browser and assist for video mining and retrieval, a novel method, video abstract, is proposed for automatically summarizing news video content. A video abstract is a compact representation of video content. It is defined to be a sequence of moving images, extracted from a longer video, much shorter than the original, and preserving the essential message of the original [16]. Content-based video abstract is an important type of content-based video mining tool. To concisely and informatively summarize news video content to maximum extent, key-frame-based abstract is put forward in this paper.

4.1 Key Frame Extraction Based on Caption and Visual Information

Key frame is the frame which can represent the salient content of the shot and summarizes contents of a video sequence. Key frame selection is an important method of summarizing a long video program. Key frames are often arranged as storyboards, in which the key frames represent shots and sequences to summarize the story. Depending on the content complexity of the shot, one or more key frame can be extracted from a single shot. In the news video, frame with caption or text is usually contain the main idea of a news story. Therefore, an algorithm based on caption and image information is introduced to extract key frame from news video. The detailed algorithm can be referred in [14].

4.2 News Video Abstract

Based on the extracted key frames, a news video abstract prototype [14] is developed using DirectShow architecture and COM criterion. Users can use this system not only to analyze news video, but also to mine and browse interested news story. Depending on user's different demand, this system can afford different searching depth. For example, you can just browse a mosaic picture of key frames firstly, and then you can choose your interested news to watch the details. Experiments reveal that news video abstract is a very effective mining tool news video database.

5 Conclusion

In this paper, we have addressed video mining techniques for efficient video organization, management and retrieval. To achieve this objective, a hierarchical video organizing model is proposed. A news video content structure mining scheme is introduced for parsing the news video into a series news stories. Both visual and audio features are real-timely extracted and utilized to analyze news video. A novel video mining tool, key-frame-based video abstract is introduced to summarize and browse the content of news video. Experimental results demonstrate the efficiency of our framework and strategies for video data mining. However, research of video mining is still primitive and much room remains for improvement. Our future work will include the video indexing techniques based on multimodal information.

References

1. U. Fayyad: Data Mining and Knowledge Discovery in Database: Implications for Scientific databases. Proceeding of Ninth International Conference on Scientific and Statistical Database Management (1997) 2-11
2. M.S. Chen, J. Han and P.S. Yu: Data mining: An Overview from a Database Perspective. IEEE Transactions on Knowledge and Data Engineering (1996), 8(6):866-883
3. J.-Y. Pan, C. Faloutsos: VideoCube: A New Tool for Video Mining and Classification, ICADL 12(2002)
4. Xingquan Zhu, Walid G. Aref, Jianping Fan, Ann Christine Catlin, Ahmed K. Elmagarmid: Medical Video Mining for Efficient Database Indexing, Management and Access. ICDE (2003) 569-580
5. Xingquan Zhu and Xiaodong Wu: Sequential Association Mining for Video Summarization. Proceedings of International Conference on Multimedia and Expo, (2003) 333-336
6. R Mohan: Text-based Search of TV News Stories. SPIE Proc. of Multimedia and Archiving Systems, NJ: SPIE Press (1996) 2-13
7. L Chen, P Faudemay: Multi-Criteria Video Segmentation for TV News. Proc. of 1st Multimedia Signal Processing Workshop, NJ: IEEE Press (1997) 319-324
8. H J Zhang, S Y Tan, S W Smoliar: Automatic Parsing and Indexing of News Video. Multimedia Systems, 1995, (2): 256-266.
9. Hauptmann A, Witbrock M.: Story Segmentation and Detection of Commercials in Broadcast News Video. http://www.ieee.org/ieeexplore, October 2000.
10. A G Hauptman, M Smith: Text, Speech and Vision for Video Segmentation: The Informedia Project. Working Notes of IJCAL Workshop on Intelligent Multimedia Information Retrieval, NJ: IEEE Press (1995) 17-22
11. J Huang, Z Liu, Y Wang, Y Chen: Integration of Multimedia Features for Video Classification Based on HMM. http://vision.poly.edu, October 2000.
12. Z Liu, Q Huang, A Rosenberg: Automated Generation of News Content Hierarchy by Intetrating Audio, Video, and Text Information. ICASSP-1999, NJ: IEEE Press (1999) 3025-3028
13. J Huang, Z Liu, Y Wang: Integration of Audio and Visual Information for Content-based Video Segmentation and Classification. Journal of VLSI Signal Processing System for Signal, Image, and Video Technology (1998), 20(2): 61-79.
14. Yu Junqing: Research of Content-Based Video Abstract. Wuhan: Wuhan University (2002)
15. Wang Weiqiang, Gao Wen: Automatic Parsing of News Video Using Multimodal Analysis. Journal of Software (2001), 12(9): 1271-1278
16. Rainer Lienhart, Silvia Pfeiffer and Wolfgang Effelsberg: Video Abstracting. Communications of the ACM (1997), 40(12): 55~62