WPM P1.10

An MPEG-4/7 based Architecture for Analyzing and Retrieving News Video Programs

Jin-Hau Kuo, Chia-Chiang Ho, M-J Shieh, J-H Yeh, and Ja-Ling Wu, *Senior Member, IEEE*
Communications and Multimedia Lab., National Taiwan University, Taipei, Taiwan

***Abstract*—In this paper, an MPEG-4/7 based architecture is proposed for analyzing and retrieving news video programs. The proposed architecture covers all stages of a content flow, such as creation, composition and presentation.**

I. INTRODUCTION

Along with the rapid success and growth of the Internet, development of broadband technology and the steady growth of computing power, the roles of the content provider have shifted from mass media to general-information end users. On account of this phenomenon, nowadays, there exists an unprecedented variety of multimedia information. In academic research, some specific kinds of video data are investigated popularly, such as sport, movie, medical and news videos. Among all the domain-specific applications, news video program processing is probably the most extensively studied topic [1]. In this paper, an MPEG-4/7 based architecture is proposed for analyzing and retrieving news video programs. The proposed architecture covers all stages of a content flow, such as creation, composition and presentation (c.f. Fig. 1).

II. THE NEWS PROGRAM VIDEO ANALYZER

The functionality of video analyzer is to generate the atomic and structural metadata for each news program clip.

A. Shot change detection:

A shot change is defined as an image content switch between two consecutive frames with different scenes. To tackle it, we propose a multi-modal-feature based shot and scene change detection algorithm (which can be viewed as a mid-stage solution between the single-modal-feature- and the semantic-based approaches). An interactive shot and scene analysis system, including dissolve change detections, is realized on the basis of the algorithm (c.f. Fig. 2).

B. News Item detection:

The semantic basic unit of a news program should be a news item. We developed a hierarchical video news program analyzer preliminary. According to this property of news item, we can regard the first frame with an anchorperson as the beginning of a news item, and a semi-automatic news video program analyzer can then be developed. After the low level shot change detection is done, all boundaries between consecutive camera shots are detected and represented as key-frames. An interactive interface is designed and user can choose which frame is with the anchorperson. The chosen frame is described by two color descriptors: color layout and color histogram. By comparing bin-to-bin color layout or color histogram distances between the anchorperson-frame and detected shot change frames (key-frames), the whole news program video can be divided into some news items of semantic meaning.

III. THE NEWS PROGRAM VIDEO COMPOSITOR

The functionality of the video compositor is to generate the rich-media and condensed units for each news program clip, respectively.

A. Re-manufacturer:

For the convenience of constructing an edit or a retrieval system, it is necessary to decompose the whole news video program into individual news item clips. For different usages, we should generate semantic video, audio and Mux units. For instance, for a speech-based recognition/retrieval system, we should generate the semantic audio unit in favor of analyzing. By taking down the temporal position (frame number) of news item, the system can find the desired clip in the GOP header and retrieve it (skipping the unwanted portion by comparing the frame number and GOP header).

B. Synthesizer:

The functionality here is to integrate heterogeneous multimedia objects, includes the video, audio and Mux units generated in the re-manufacturer module. Its purpose is for those rich media consumption, such as powerful PCs and interactive TV. Our proposed system integrates MPEG-4 scene description framework, JavaScript language, and visual editing environment to build a RAD tool for multimedia. We named it RMMAD (Rapid Multimedia Application Development) system. This RMMAD tool provides PME (Property-Method-Event) programming style to facilitate the development of a multimedia application (c.f. Fig. 3).

IV. THE NEWS PROGRAM VIDEO PROFILER

In the MPEG-7, a lot of visual, audio and multimedia description interfaces have been proposed and standardized [2,3,4]. But these content description tools seem too general to be adapted to a specific application. In the followings, we identify the notations appropriate to the news program videos.

- Video Meta notations
  - Date (ex. 2002/1/12 afternoon), Location, Interviewer (ex. TVBS Lee Tao)
  - Roles (ex. Legislative Yuan Speaker Wang Chin-ping)
  - Time code (ex. 00:59:28), Duration (ex. 1'10")
  - Media Type (ex. Betacam-VX)

0-7803-7721-4/03 $17.00 © 2003 IEEE

- Category (ex. Politics)
- Event description (ex. Election for Taipei city mayors)
- Textual information

- Video Data notations
  - Color (ex. Color layout), Motion (ex. High motion)
  - Camera motion (ex. Zooming), Caption (ex. Headline)
  - Audio (ex. The voice of anchorwoman)
  - Segment structure (ex. Shot, scene…)
  - Object (ex. Anchorwoman)
  - Region of interest (ex. Position of roles)
  - Preview (ex. Summarization), Linkage to Copyright

V. THE NEWS PROGRAM VIDEO BROWSER

Two kinds of applications are investigated for thin and rich clients, respectively. They can be regarded as pull and push applications described in MPEG-7 standard.

A. Speech Recognition Retriever:

We realized a MPEG-4 FGS decoder for constrained embedded systems [5]. Thus, users can query the news item they want by using a speech-based recognition/retrieval system, and can view the streamed news video clips from a PDA (c.f. Fig. 4). The retriever has been tested by using one-week life news video program taken from one of the local TV stations in Taiwan (approximately two hours per day). The preliminary test result is rather promising (in which the investigators includes both the TV station's administrators and members of our Laboratory).

B. Interactive Player:

Providing VoD services on next generation TV is demanded by the current market. MPEG-4 seems to be the key technology to make it possible. With MPEG-4, the content provider may give the indices of hot news for user selections, or offer additional instant financial information simultaneously. On the other side, users may customize the layout or its presentation style. To separate the control and display, multiple scenes supported in MPEG-4 will enable control elements to be centralized on the display of the remote control device such as the LED on a PDA (c.f. Fig. 5).

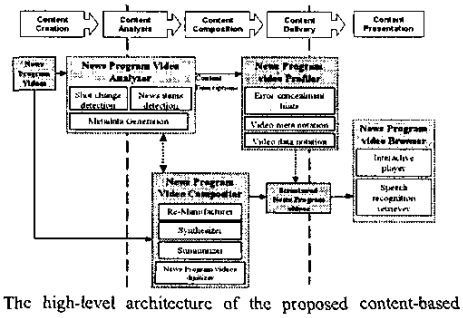

Fig. 1: The high-level architecture of the proposed content-based news program videos processing framework.

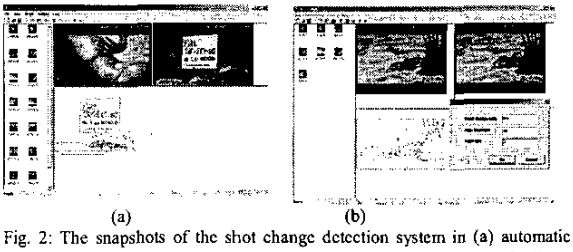

Fig. 2: The snapshots of the shot change detection system in (a) automatic mode, and (b) semi-automatic mode, with the user-interactive feedback interface.

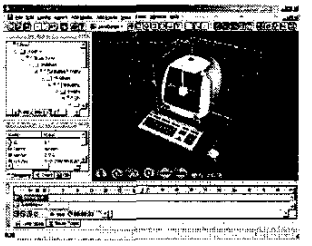

Fig. 3: The appearance of the RMMAD system.

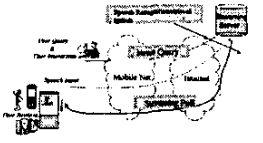

Fig. 4. The blocks diagram of a news video browser.

Fig. 5: News Report and Video On-Demand for TV Applications: In this example, a pocket controller appeared on the left-down corner controls the layout of TV according to users' preference. Four lienable informative objects are presented: a list of headline news for user selection (left), information about selected headline (down), virtual reporter using text-to-speech technology (right-down), and the movable stock information to deliver update data (right-up). Users may customize the layout style by loading appreciative scene-stream from the server or from local storage.

REFERENCE

[1] S. Boykin and A. Merlino, "Improving broadcast news segmentation processing," in Proc. IEEE Int. Conf. Multimedia Computing and Systems, vol. 1, 1999, pp. 744–749.

[2] ISO/IEC 15938-3: Information technology -- Multimedia content description interface – Part 3: Visual, 2002.

[3] ISO/IEC 15938-4: Information technology -- Multimedia content description interface – Part 4: Audio, 2002.

[4] ISO/IEC JTC1/SC29/WG11, Coding of Moving Pictures and Audio, N3966, Part 5 Multimedia Description Schemes.

[5] H-K Tsao and J-L Wu, "An Efficient Macroblock-Sync Decoding Architecture for MPEG-4 FGS Video," in 45th IEEE International Midwest Symposium on Circuits and Systems.

[6] Y.-S. Tung, and J.-L. Wu, "Architecture Design, System Implementation, and Applications of MPEG-4 Systems," Proc. of 2nd Workshop and Exhibition on MPEG-4, pp.37-40, 2001.