ELSEVIER

Constructing and application of multimedia TV-news archives ☆

H.T. Pao [a,*], Y.H. Chen [b], P.S. Lai [b], Y.Y. Xu [b], Hsin-Chia Fu [b]

[a] *Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan, ROC*
[b] *Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan, ROC*

Abstract

This paper addresses an integrated information mining techniques for broadcasting TV-news. This utilizes technique from the fields of acoustic, image, and video analysis, for information on news story title, newsman and scene identification. The goal is to construct a compact yet meaningful abstraction of broadcast TV-news, allowing users to browse through large amounts of data in a non-linear fashion with flexibility and efficiency. By adding acoustic analysis, a news program can be partitioned into news and commercial clips, with 90% accuracy on a data set of 400 h TV-news recorded off the air from July 2005 to August 2006. By applying speaker identification and/ or image detection techniques, each news stories can be segmented with a better accuracy of 95.92%. On-screen captions or subtitles are recognized by OCR techniques to produce the text title of each news stories. The extracted title words can be used to link or to navigate more related news contents on the WWW. In cooperation with facial and scene analysis and recognition techniques, OCR results can provide users with multimodal query on specific news stories. Some experimental results are presented and discussed for the system reliability, performance evaluation and comparison.
© 2007 Published by Elsevier Ltd.

Keywords: TV-news archives; Multimedia; Information mining; Multimodal query; Video OCR

1. Introduction

Among the major sources of news program, TV has clearly had the dominant influence atleast since the 1960s. Yet it is easy to find the old newspaper in microfilm in any public library, but it is impossible to find the old footage of television news in the same library. TV-news archive has existed in the United States for 35 years. Paul C. Simpson founded the Vanderbilt University Television News archive in 1968. In Huffman, Yang, Yan, and Sanders (1990), a team in the University of Missouri-Columbia decided to do a content analysis of the three US network coverage of the 1989 Tiananmen Massacre, they located these news items in the Vanderbilt Archive Index Vanderbilt television. The Vanderbilt archive promptly provided the 11 h-video clips all related to the Tiananmen Massacre.

At the same time, the Missourian team also planned to do a comparable study of Taiwanese reportage on Tiananmen Massacre. But the equivalent material of the Vanderbilt archive did not exist in Taiwan then. Therefore, that study only contained the US perspective of the Tiananmen Massacre. This paper proposes an integrated methodology for the information mining on a multimedia TV-news archive in Taiwan. As described in Xu, Chen, Tseng, Lai, and Fu (2004) Lai, Lai, Tseng, Chen, and Fu (2004), a fully automated Web-Based TV-news system were implemented to achieve the following goals:

1. *Academic and applied aspects*: This archive will greatly improve the quality of TV-news. As Dan Rather, the CBS anchorman, once mentioned that he lives with two burdens – *the ratings and the Vanderbilt Television News Archive* Therefore, once the archive is there, the researchers and the public will do some content analysis on the TV-news. And the journalists will be more careful in what they report.

2. *Timing factor*: Vanderbilt archive started its project with Betacam videotapes in 1968. There will be a problem of preservation because these tapes deteriorate along the years. Today, we can save all the TV-news in hard disc, VCD or DVD.

Informedia (http://www.informedia.cs.cmu.edu/) is an integrated project launched in Carnegie Mellon university. Its overall goal is to use modern AI techniques to archive video and film media. VACE-II Informedia (http://www.informedia.cs.cmu.edu/arda/vaceii.htm), a sub-project of Informedia, automatically detects, extracts, and edits highly interested people, patterns, and story evolves and trends in visual content from news video.

This paper proposes a integrated information mining technique that can automatically generate semantic labels from news video Daniel and Daniel (2002), and statistical methods to discover hidden information. We intend to expect that the following significance will come to exist.

- Although web-news provides another efficient way to access news, watching TV-news already became habit of many people. Beside this, most of web-news system can only provide text-based news.
- There are so many channels providing TV-news. People need more information for searching like-minded channel.
- Although almost every channel announced that they are dispassion, real dispassion is hard to archive with human editing. We need some evaluation to check if the channel is really dispassion.

The rest of this paper is organized as follows. Section 2 introduces the overall concepts of the multimedia TV-news archive. In Section 3, methods of generating necessary semantic labels from the recording TV-news video are presented. Section 4 focus on describing the information mining from these semantic labels. Finally, summary and concluding remarks are given in Section 5.

2. TV-news archive

A fully automated Web-based TV-news system Lai, Lai, Tseng, Chen, and Fu (2004); Xu et al. (2004) consists of three modules: (1) TV-news video acquisition and content segmentation, (2) news content analysis, and (3) user interface for news query, search and retrieval. Fig. 1 depicts the overall architecture and interaction of these three modules. The major tasks of the acquisition module are to record TV-news programs in a proper video format, and to fetch related news text contents from Internet webs. Content analysis module segments the recorded news video into story based units, and extracts news title and keywords from each story unit. Providing a friendly querying and browsing environment for retrieving interested news stories is the most important task of the user interface module.

An overall news video processing and content analyzing are depicted in Fig. 2. At the beginning, a TV-news pro-
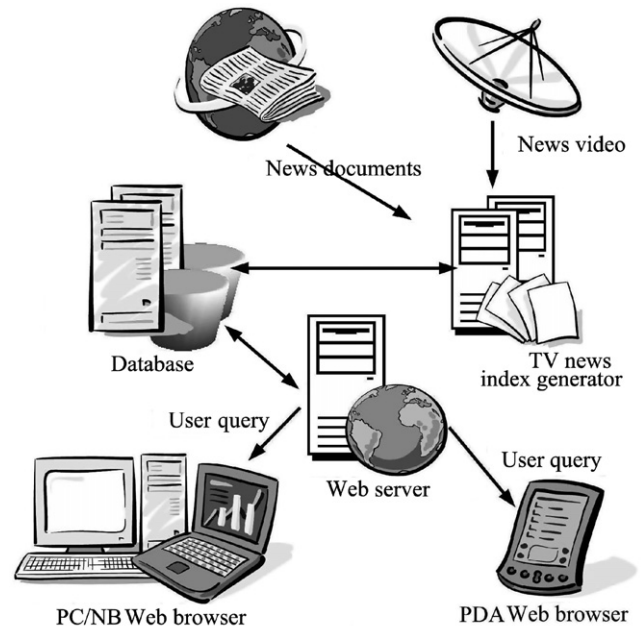

Fig. 1. The overall architecture and information processing flow of the proposed fully automated web-based TV-news system.
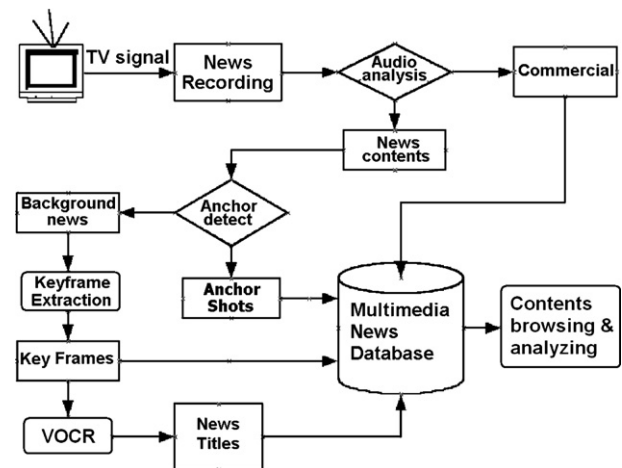

Fig. 2. The flow diagram of TV news acquisition and content segmentation.

gram is captured and encoded into stream video format (Iain & Richardson, 2003; Wang, Ostermann, & Zhang, 2002). The recorded streaming video is named with dates first, and then stored in database. In the meantime, a shot detector is used to segment the streaming video into scene based shots for news unit generation and key-frame extraction (Lee, Yoo, & Jang, 2006; Patel & Sethi, 1997). Within a scene shot, speaker identification techniques are then applied to detect anchor frames (Cheng, Wang, & Fu, 2004). The close-captions in the anchor frames are then extracted and recognized by using video OCR techniques (Lin, Liu, & Chen, 2001) as candidates for the news title and keywords of each news units. The extracted keywords can then be used to match with (1) Internet news stories to construct links between TV-news stories and Internet news,

and (2) the users' query words for retrieving interested and/or related news stories. In addition, the extracted characters by video OCR from each news units can also be used as semantic labels of the news units. More description on the semantic labels will be presented in Section 3.3.

3. News information tree generation

The most important things in news story writing are that journalists commonly refer to as the 5 W's: who, what, when, where and why. These questions are crucial for catching a reader's attention and introducing the essential facts of the story. Standing on this basic rules, the news archive system introduced in Section 2, are further improved to extract more information from a recorded news video. A *news information tree* (Feinstein & Morris, 1988) is suggested to structure the contents of recorded video clips for helping the user focus on specific news information, and information that is a little more general.

3.1. News information tree

Fig. 3 illustrates a hierarchical structure of a news information tree. The hierarchical tree contains five types of video information records: (1) date (when), (2) channel (where), (3) title (what), (4) content (how), and (5) commercial. The title record contains the starting time, length, and brief description of the corresponding video clips. The content record can also be further divided into the following sub-records: (a) on-site locations, (b) interview, and (c) tables or quoted word.

3.2. Analysis units

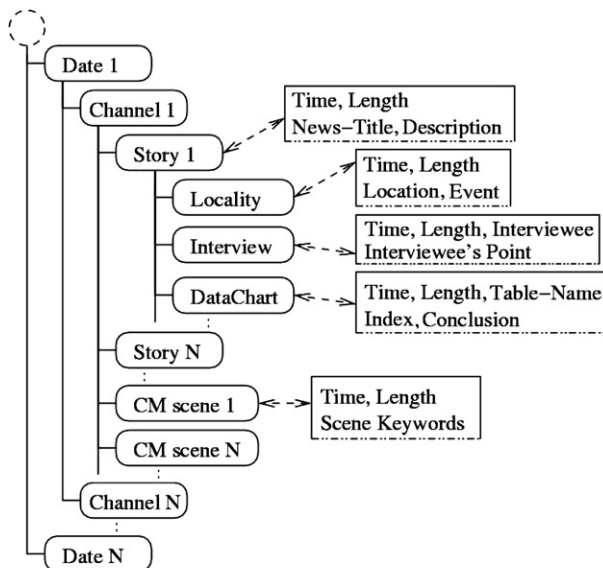Usually, a TV-news program contains the following items: news stories, commercials and weather reports.

Complete description of shot detection and scene segmentation can be found in Huang, Lai, and Fu (2004). A flow chart of TV-news program segmentation is shown in Fig. 4, and brief introduction are described as follows. Among various scene shots, anchor video clips are detected first (Kim, Kim, Ra, & Choi, 1999).

In general, anchor segments are the most appeared video clips (Saracoglu, Tutuncu, & Allahverdi, 2007), thus we propose to use BIC (Fraley & Raftery, 1998), an unsupervised method, to cluster anchor segments from the other clips. As shown in Fig. 5, this method contains the following procedures:

1. The MFCC audio feature sequence X is generated from input audio at first.
2. BIC segments X into segments $X_1, X_2, \ldots X_n$.
3. These segments then are clustered as several clusters $C_1, C_2, \ldots, C_3$.
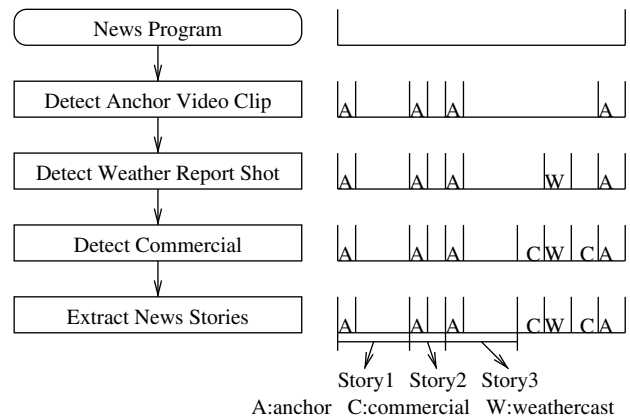4. The cluster containing most clip segments is the set of anchor clip.

Fig. 4. The flow diagram of the proposed news story analysis and information extraction processes.

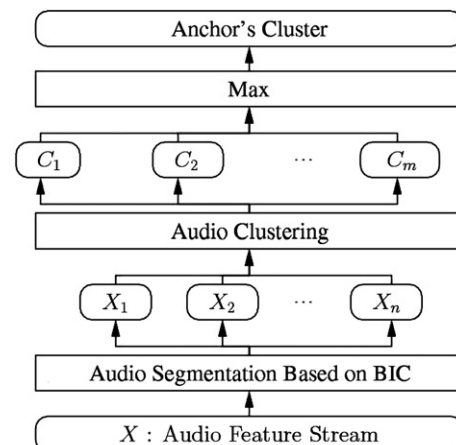Fig. 3. The data structure of a news information tree.

Fig. 5. The flow diagram of a BIC-based audio segmentation method.

After locating each anchor shot, a SVM model based video classifier (Sun, Tseng, Chen, Chuang, & Fu, 2004) is used to detect weather report shots. Finally, commercials are detected and separated from on-site news stories. The following feature detecting techniques are integrated to achieve a high performance commercial detector:

- The variation rate of zero crossing rate.
- Short time energy.
- Shot change rate.
- Clip length.

At this stage, anchor's briefing, weather reports, commercial, and the background stories are all separated and identified. For each on-site stories, as shown in Fig. 6, we further segment and classify each on-site scene into three categories: locations, interview and tables or quoted words (what).

Fig. 7 shows how to partition the on-site news story into (1) location, interview, and tables or quoted words scenes. In general, on-site narration is not active during the interview scene, as shown in Fig. 6, the interview scene can be distinguished from location scene. By using its special char-

acteristics of the character regions, tables or quoted words scene can be distinguished from location scenes.

3.3. Semantic labels of units

This section describes how to assign each segmented unit with semantic labels. Basically, the text words for each label are extracting from text streams in close-caption.

Usually, a TV-news program often provides audience a quick overview of each news story in on-screen captions, such as names of location, people, and keywords of events, ... etc. In general, these texts are quiet enough to give enough information for labeling each segmented units.

Fig. 8 shows how to establish a news information tree. The information tree establishing process contains two phases: story and scene phases.

In story phase, all on-screen characters are recognized by video OCR first. Then, the recognized characters or words are used to match with text-based news documents, which are usually retrieved from Internet. The title and contents of best matched text-news document not only fill out the story information record of news information tree, but also used to picking label candidates, including locations, people names, event words, quoted phrases, and tabular data, up for scene phase processing.
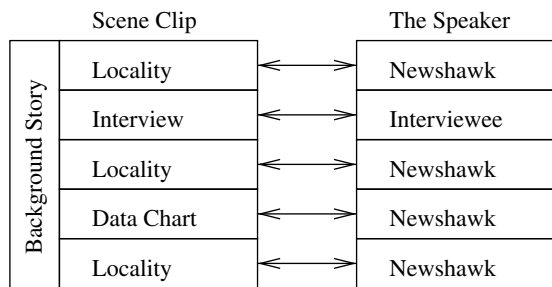

Fig. 6. The general structure of a news story. On-site scene story contains three major news contents: locations, interview and tables or quoted words.
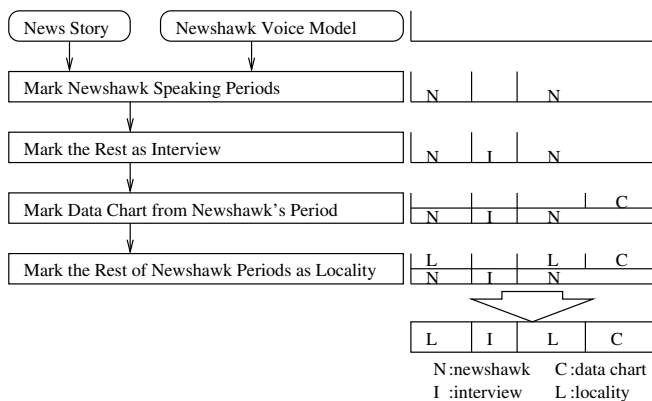

Fig. 7. On-site scene segmentation flow. The narration periods can be detected by using speaker identification techniques to distinguish narrator's speech voice from the rest of scenes. Detecting the screen characters regions can find the tables or quoted words scenes. Then, the rest of scenes that are not belonging to interview or tables or quoted words scenes must be the location scenes.
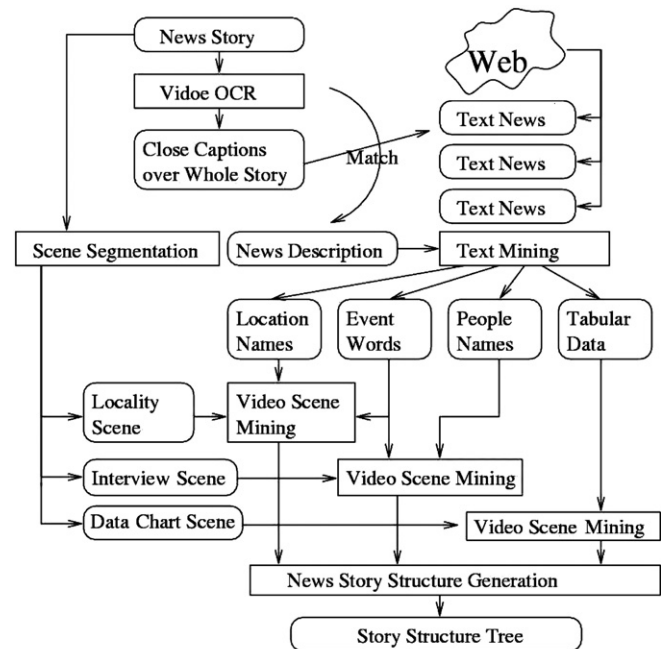

Fig. 8. Information flow of the generation of a news information tree. For each story, video OCR (optical character recognition) technique is applied to extract characters in the close-caption. The extracted characters will then be used to match with the retrieved news document over the Internet web sites. From the matched documents, key information, such as associated people, event location, reporters' names etc., can be retrieved accordingly. Finally, video clips associated with location, interview, tables and quoted words scene can then be labeled according to the extracted keywords.

In scene phase, picking semantic labels up from label candidates for each scene is done in this phase. As shown as Fig. 9, the on-screen captions of locality scene provide location name and event descriptions. Therefore, in locality scene, the location and event words of label candidates are searched from on-screen captions to find which ones exactly appeared.

Fig. 10 is an example of interview scene. In interview scene, the interviewee's name and their points are always given by on-screen captions. Therefore, in interview scene, we search people name, and events word of label candidates instead.

The example of data chart scene is shown as Fig. 11. In general, on-screen captions fill data chart scene. These captions may present quoted sentence or tabular data. There-

Fig. 11. An example of data chart scene frame. The data chart scene is used to present information in a organized manner. Additional information is also available from the on-screen characters.

fore, searching for quoted sentence or tabular data in data chart scene is the major task.

4. Data mining on the news information tree

This section presents how and what to mine from a news information tree (NIT). In Section 4.1, we propose to mine the favored or preferred news contents of a TV-station. The news information tree can also be used to track the evolution of a series of news stories (see Section 4.2). In addition, the mining results from the NIT and the realtime ratings can be combined to provide TV-news commercial buyers a very useful guidance.

4.1. Mine the news preference of a TV-Station

Generally speaking, a TV-station arranges the broadcasting sequence of each story in a news program according to their impact and attractiveness to audience. In fact, a preferred news story often gets more time on the air. By analyzing the sequence order and the length of stories, the preferred or the favored news stories of a TV-station can be roughly estimated or judged. Mining the NIT to extract favored or preferred types of news story from a TV-station will help audience to find the favored news channel.

The proposed news mining method is described as follows: Given N sets of keywords, $K_1, K_2, \ldots, K_i, \ldots, K_N$, which correspond to N news topics (or subjects), let the following delta function $\delta(k, K_i)$ define the relations between a keyword k and a keyword set K_i:

$$\delta(k, K_i) = \begin{cases} 1, & \text{if keyword } k \in K_i \\ 0, & \text{otherwise.} \end{cases}$$

1. Extract keywords $\{k_{s_j}^l; l = 1, \cdots, L_j\}$ from a scene unit s_j in a news program.
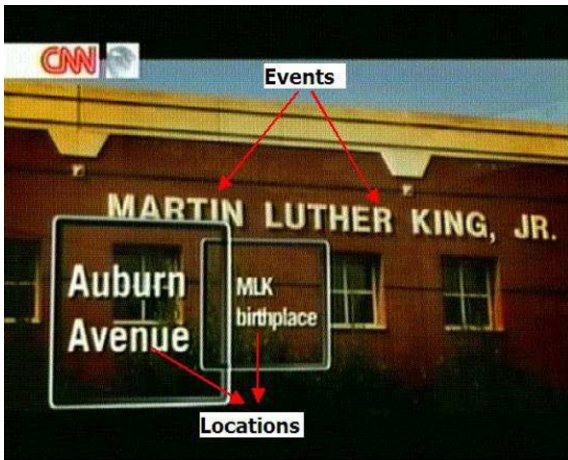

Fig. 9. An example of locality scene frame. The locality scene is used to show where and what news occurred. Thus, the location information and event description can be retrieved from the close-captions of a locality scene.

Fig. 10. An example of interview scene frame. An interview scene is used to present a reporter's point of view. Normally, the reporter's name and opinions can be extracted from subtitles or closed captioning.

2. For each scene units s_j, compute its association frequency $F(K_i|s_j)$ with respect to a subject K_i,

$$F(K_i|s_j) = \frac{\sum_{l=1}^{L_j} \delta(k_{s_j}^l, K_i)}{\sum_{i=1}^{N} \sum_{l=1}^{L_j} \delta(k_{s_j}^l, K_i)}$$

3. Compute the the association frequency of news program $F_d(t|K_i)$ at time t and day d:

$$F_d(t, K_i) = \begin{cases} F(K_i, s_1), & \text{for } s_1.\text{start} \leqslant t \leqslant s_1.\text{end} \\ F(K_i, s_2), & \text{for } s_2.\text{start} \leqslant t \leqslant s_2.\text{end} \\ \vdots & \vdots \\ F(K_i, s_M), & \text{for } s_M.\text{start} \leqslant t \leqslant s_M.\text{end}, \end{cases}$$

where $s_j.\text{start}$ and $s_j.\text{end}$ are the start and the end time of a scene unit s_j in a news program at time t of the day d.

The associated frequency distribution from one segment of news program is not enough to represent the overall preference or trend of a news channel, thus long term statistics is needed.

By accumulating a longer period (say one month) of associated frequency of news subjects, the preference of a channel can be discovered. As shown in Fig. 12, keywords that are related to social news, political news, and entertainment news are applied to associate with and to accumulate frequency of news topics. As we can see in this example, the monitored news channel favors social and political news more than entertainment news.

4.2. The evolution of a series of news stories

The evolution of a news story can also be mined from the news information tree. By associating the keywords of a specific event with recorded news scenes over a period of days, then the accumulated association frequency of matched scene units presents an overall developing and progressing of the specific news stories. Fig. 13 shows a sort of life-cycle of a particular news events. In addition,
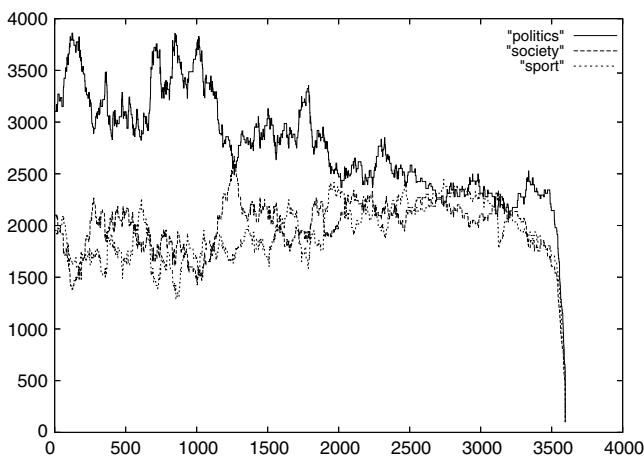

Fig. 12. Three sets of (representative) keywords are used to associate the appearing frequency of *social*, *political* and *sport* news in a news program.
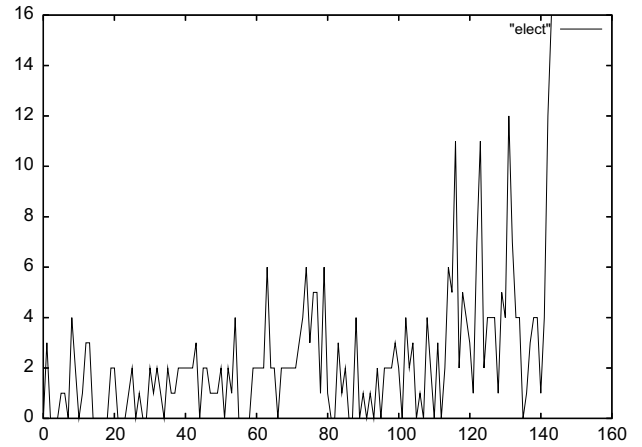

Fig. 13. The life-cycle of a specific news events along with a period of time.

the spreading of the specific events to other areas, e.g. cities, counties, countries, etc., can also be retrieved from the associated names of locations in the matched scene units. For example, one can query a news story by using a particular people's name, then the person's daily schedule and/or whereabouts can be retrieved from the recorded NIT.

4.3. The mining on TV commercial

Beside background stories, commercial records are also valuable information. Huang et al. (2004) proposed commercial detecting and identifying methods in TV video clips. When a commercial frame contains image keywords in a video frame, video OCR techniques can be used to extract keywords to label the corresponding video clips. Otherwise, keyblock-based image retrieval methods (Zhu, Rao, & Zhang, 2002) may be utilized to represent and to identify each commercial clips. However, manual annotation is needed to label the keyblock. By gathering statistical information of these labels and keywords in news programs, cross relationship between TV commercials, real-time ratings, and news stories can be observed and analyzed to achieve a useful marketing database. Two example areas, *customer modeling* and *cross-selling*, in database marketing are discussed in the following.

4.3.1. Customer modeling

The basic idea behind customer (i.e. the commercial buyers and news audience) modeling is to improve audience response rates by targeting prospects that are predicted as most likely to respond to a particular advertisement or promotion. This is achieved by building a model to predict the likelihood that groups of news audience will respond based on news type, viewing time and news channels as well as previous viewing behavior. In addition, by targeting more effectively to prospects and existing commercial buyers, TV-station operators can improve and strengthen customer relationships. The customer can perceive more value in TV-

news and commercials (i.e. both commercial buyers and news audience receive only products and/or services of interest to them).

4.3.2. Cross-selling

The basic idea behind cross-selling is to leverage the existing customer base by selling them additional products (commercial time slots) and/or news services. By analyzing the groups of products or services that are commonly purchased together and predicting each customer's affinity towards different products using historical data, a TV-station can maximize its selling potential to the existing customers. Cross-selling is one of the important areas in database marketing where predictive data mining techniques can be successfully applied. Using historical purchase data of different products from the customer database along with news type, viewing time and news channels, commercial buyers can identify their products that are most likely to be of interest to targeted news audience. Similarly, for each type of product (i.e. commercial or groups of commercials), a ranked list of different types of news or groups of audience, that are most likely to be attracted to that product. Then, arrangement of commercials with matched types of news to achieve a high likelihood of audience response rate.

5. Conclusion

This paper addresses techniques and possible applications of fully automated information mining on a multimedia TV-news archive. The proposed automated information mining contain the following processes: (1) segmenting a TV-news program video recording into scene clips, (2) using video OCR to extract and recognize close-caption and/or image characters into keywords for each scenes, (3) using keywords to generate semantic labels for each scenes, and (4) segmenting commercial video clips from news clips. Information associated with various labels and scenes (e.g. the starting and ending time of a scene) are stored in the proposed *news information tree*. Performing statistical analysis on the data items in the news information tree can reveal hidden information, like popular channels and evolution of some hot news stories. These information can help general multitude in finding their favored or desired news channel, searching focal point person, tracking hot news stories, ..., and so on.

References

Cheng, S.-S., Wang, H. m., & Fu, H.-C. (2004). A model-selection-based self-splitting gaussian mixture learning with application to speaker identification. *EURASIP Journal on Applied Signal Processing, 17*, 2626–2639.

Daniel, Gildea, & Daniel, Jurafsky (2002). Automatic labeling of semantic roles. *Computational Linguistics, 28*(3), 245–288.

Feinstein, C. D., & Morris, P. A. (1988). Information tree: A model of information flow in complex organizations. *Systems, Man and Cybernetics, IEEE Transactions, 18*(3), 390–401.

Fraley, C., & Raftery, A. E. (1998). How many clusters? which clustering method? answers via model-based cluster analysis. *Computer Journal, 41*, 578–588.

Huang,T. -Y., Lai, P.- S., & Fu, H.-C. (2004). A shot-based video clip search method. In *Proceedings of CVGIP2004*, Taipei, Hualien, ROC, August 2004.

Huffman, S., Yang, T., Yan, L., & Sanders, K. (1990). Genie out of the bottle: Three US Networks report tiananmen square. In *Proceedings of the annual meeting of association for education in journalism and mass communication*, Minneapolis, Minnesota, USA.

Iain, E. G., & Richardson, H. (2003). *264 and mpeg-4 video compression.* Wiley Press.

Kim, D.-W., Kim, J.-T., Ra, I.-H., & Choi, Y.-S. (1999). A new video interpolation technique based on motion-adaptive subsampling. *IEEE Transactions on Consumer Electronics, 45*(3), 782–787.

Lai, P. S., Lai, L. Y., Tseng, T. C., Chen, Y. H., & Fu, H. C. (2004). A fully automated web-based TV-news system. In *Proceedings of PCM 2004*, Tokyo, Japan, Dec. 2004.

Lee, M. H., Yoo, H. W., & Jang, D. S. (2006). Video scene change detection using neural network: Improved art2. *Expert System with Applications, 31*(1), 13–25.

Lin, C.-J., Liu, C.-C., & Chen, H.-H. (2001). A simple method for chinese video ocr and its application to question answering. *International Journal of Computational Linguistics and Chinese Language Processing, 6*(2), 11–30.

Patel, N. V., & Sethi, I. K. (1997). Video shot detection and characterization for video databases. *Pattern Recognition, 30*(4), 583–592.

Saracoglu, R., Tutuncu, K., & Allahverdi, N. (2007). A fuzzy clustering approach for finding similar documents using a novel similarity measure. *Expert System with Applications, 33*(3), 600–605.

Sun, S.- Y., Tseng, C. L., Chen, Y. H., Chuang, S. C., & Fu, H. C. (2004). Cluster-based support vector machine in text-independent speaker identification. In *Proceedings of international joint conference on neural networks IJCNN* 2004, Budapest, Hungary; 2004.

Vanderbilt television news archive, http://www.vanderbilt.edu/vtna.

Wang, Y., Ostermann, J., & Zhang, Y.-Q. (2002). *Video processing and communications.* Prentice Hall Press.

Xu, Y. Y., Chen, Y. H., Tseng, C. L., Lai, P. S., & Fu, H. C. (2004). Multimedia TV-news browsing system. *In Proceedings of IEEE international conference on multimedia and expo (ICME)*, Taipei, Taiwan, ROC; June 2004.

Zhu, L., Rao, A., & Zhang, A. (2002). Theory of keyblock-based image reieval. *ACM Transactions on Information Systems, 20*(2), 224–257.