An Internet News Video Browsing System*

Shih-Che Lin, Yeong-Yuh xu, and Hsin-Chia Fu
The Department of Computer Science and Information Engineering
National Chiao-Tung University, Taiwan
hcfu@csie.nctu.edu.tw

H.T. Pao
The Institute of Information Management
National Chiao-Tung University, Taiwan

Abstract

With the amount of multimedia information growing and limitation of network bandwidth, it is necessary to construct a system which helps people to filter every kind of multimedia information in the network automatically. Due to the large amount and the linearly access property of video data, we always need a convenient and efficient tool for video browsing and indexing to determine if the video data is interesting. And we aim at the most important video domain, News program, to parse and segment. After reconstructing the News, we try to provide as many as possible browsing ways for users. In this paper, we bring up an idea to separate commercial shots from the News program and to detect anchor shots. We describe algorithms for constructing News stories with those classed shots. At the same time, we describe a whole framework of News video browsing system, including the flow path of system, design of database, and the transportation of video in the network.

1 Introduction

It's essential condition for people to receive news anytime. So traditional newspapers set up electric newspaper webs to provide firsthand news. Because an electric newspaper is short of video and audio, it can't express news clearly. By the way, the TV news programs which play all day lack interactive interface and are limited to broadcast only, so users can't query or watch the news that they are interested in.

Therefore, we establish the News Video Browsing System(NVBS) to merge the advantages of electric newspapers and TV news programs[1][2]. The functions of NVBS is:

1. to provide users to query news by dates and channels;
2. to provide users every kind of way to browse news;
3. to provide users to query news by keywords.

In order that, we have to build up the content hierarchy of news programs first. The content hierarchy of news programs is shown in Figure 1[3]. That is, we should divide news from commercials, then construct all stories from news and segment them into an anchor shot and detailed news reporting(non-anchor shots).

We also design database to store the processed news information and build a friendly interface to let users send request or receive messages easily.

2 Structure and Flow Path of NVBS

The NVBS includes three subparts:

1. automatic parsing and constructing news video story system;
2. video segmentation and summary in news video[1];
3. news video content search system[2].

In this paper, we will introduce the integration and flow path of NVBS, and the automatic parsing and constructing news video story system.

2.1 Flow Path of NVBS

The flow path of NVBS is shown in Figure 2. We describe it as follow:

1. convert the NTSC signal into digital AVI. format by capture card;

*This research was supported in part by the National Science Council under Grant NSC 89-2213-E-009-094.

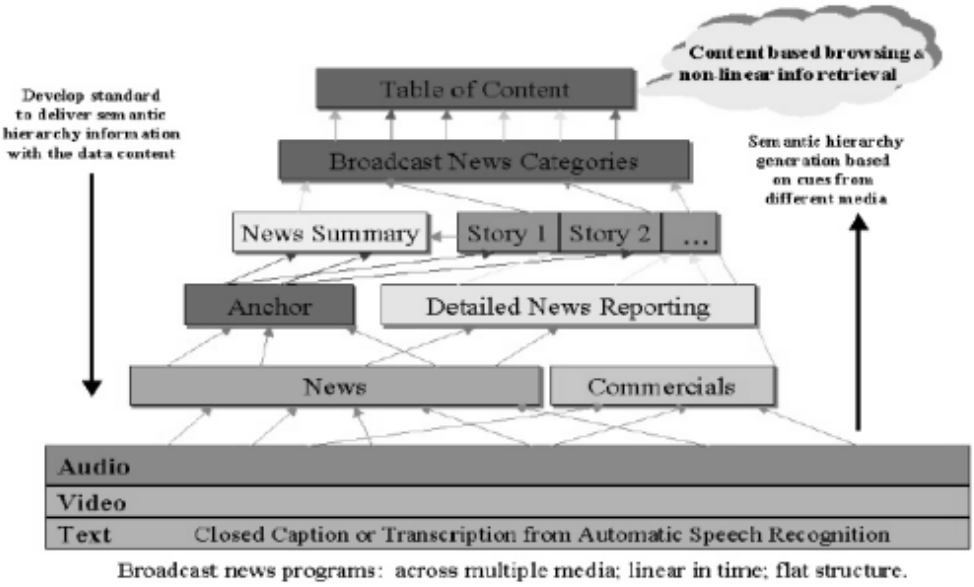

Figure 1. The content hierarchy of news programs

2. extract audio from AVI. and encode the AVI. into ASF. format;

3. separate the news program into news and commercials by audio processing;

4. dump the ASF. to extract JPEG frames;

5. process news in video by 3 and 4;

6. find out the anchors shots and construct the content hierarchy of the news program by 5;

7. make a summary for every news story[1];

8. carry out news video content query and search system by 5[2];

9. store all meaningful information into database;

10. send correct ASF. file to client by user query request.

2.2 Database Design

We need to design database to store the information of news and employ a standard database language–SQL. At the same time, we design three tables to satisfy kinds of demands:

1. Program table: Every row in table represents a news program, including 5 columns that're date_ID, date, channel, name, and filepath.

2. Commercial table: Every row in table represents a commercial segment, including 5 columns that're commerical_ID, date_ID, order, start_time, and end_time.

3. Story table: Every row in table represents a story, including 7 columns that're story_ID, date_ID, order, start_time, medium_time, end_time, and title.

2.3 Transportation of Video

In NVBS, news video is stored in ASF(Advanced Streaming Format) and transported to client through Windows Media Services. Then we can control which segment of news video is played by ASX(ASF Stream Redirector) language[9].

3 Separation of News and Commercial

For an original news program, we have to separate it into news and commercials first. After this initial process, we can continue the following job, building stories. At the same time, it can reduce the computation and complexity of the following job and raise the correctness of system.

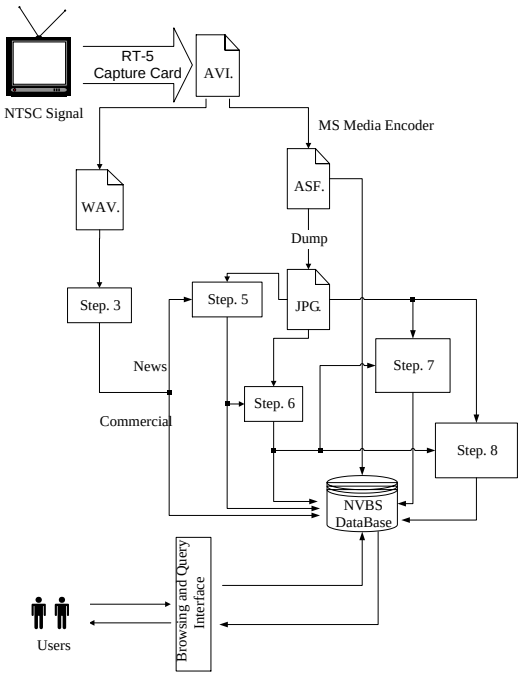

Figure 2. The flow path of NVBS

3.1 Audio Process Applied to Commercial Detection in News Video

In the past, the audio of commercials usually belongs to music, and the audio of news belongs to speech. Then we can separate news from commercials by the difference between speech and news[3][5].

Therefore, commercial ideas are more and more originally now. To follow the tread, the previous method for detecting news and commercials couldn't work well anymore. After observing, we discovered that it's hard to divide news from commercials in video domain. On the other hand, we can find out that there is a segment of special music between news and commercials in audio domain. In other words, if we could find out the special music in news programs, the boundary between news and commercials is found. It is shown in Figure 3.

Figure 3. manifestation of news and commercials in video and audio domain

The method for searching special music is as follows,

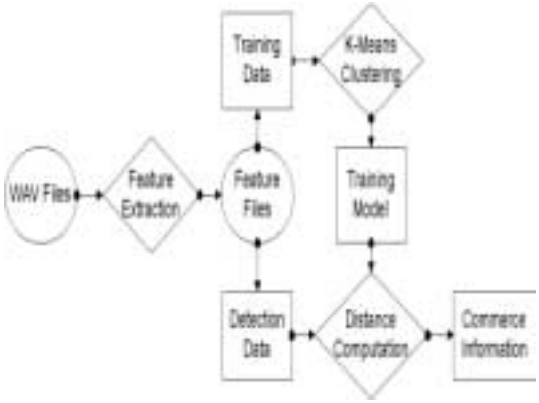

Figure 4. Flow path of separation of news and commercials in audio

extracting the feature of source WAV file in the spectrum of frequency first, and using k-means clustering to make a training model in advance. After comparing the training model and testing data, we can find out where is special music. In the end, we can separate commercials from news by these segments of special music. It's shown in Figure 4.

3.2 Feature Extraction

Because the change of audio in frequency is slower, we translate source into frequency domain to extract the feature[6][7]. Then our sampling rate is 11KHz, and the length of each frame is 20ms, and the distance between adjacent frames is 10ms.

3.3 Build Training Model

We select special music to be training data, and then train it to be training model. We use k-means clustering to cluster training data into k groups. The k-means algorithm is:

Step 1. Select random k points to be centers of k clusters;

Step 2. Assign each point to the cluster whose center is the closest;

Step 3. Update centers;

Step 4. Re-assign all points;

Step 5. If any point is moved to a different cluster, go back Step 2, else finish algorithm.

3.4 Comparison between Testing Data and Training Model

When we compute the distance between testing data and training model, we apply a concept of Gussian Mixture

Model. That is, we add 2 new parameters, Weight and Variance, to each cluster. So the equation of similarity between a point x and model U is:

$$\text{Similarity}(x, U) = \sum \frac{\text{variance}(k)}{\text{distance}(x, k)} * \text{Weight}(k) \quad (1)$$

and

$$\text{Weight}(k) = \frac{\text{\# of points in cluster k}}{\text{\# of points in model U}} \quad \text{where } k \in U \quad (2)$$

$$\text{Variance}(k) = \frac{\sqrt{\sum \text{distance}(x, k)^2}}{\text{\# of points in cluster k}} \quad \text{for all points x in cluster k} \quad (3)$$

Then we can describe models more clearly.

3.5 Construct News and Commercial Structure

As a rule, the whole structure of news and commercials is shown in Figure 4. It's news between first two special music segments, and the next is a commercial, and so on. The news and commercials usually last for certain time, so we will merge the music segments which close near too much. And a special music segment must be more than 1 second, so we destroy too short ones.

4 Anchor Detection and News Story Construction

For viewers, a meaningful news story is the set of the anchor shot and detailed news reporting(non-anchor shots). Therefore, a basic unit of a news is a anchor shot or non-anchor shots. Then two units which have the same meaning could be formed to be a news story. We bring up a method for constructing news stories automatically to let a user browse the story which he is interested in only and decide how to browse it. The way we provide to browse a news story includes (1)keywords[2], (2)keyframes[1], (3)anchor shot, (4)detailed news reporting, and (5)summary[1]. Thereupon most time and bandwidth which is wasted to access non-necessary files is saved.

4.1 Anchor Detection

After previous process, we can get pure news video without commercials. Then we need to find out anchor shots, and we decide to do it in video domain. The reason is shown in Figure 5, and (a)(b)(c) represent three kinds of news stories:

1. Fig.5(a): After the anchor summarizes the news story, the video is changed to the scene shot on location, and reporters speak aside.
2. Fig.5(b): There are anchor and non-anchor shots, but anchor's voice cover whole news story instead of reporters.
3. Fig.5(c): There is only an anchor shot without non-anchor shots, and then another story goes on.

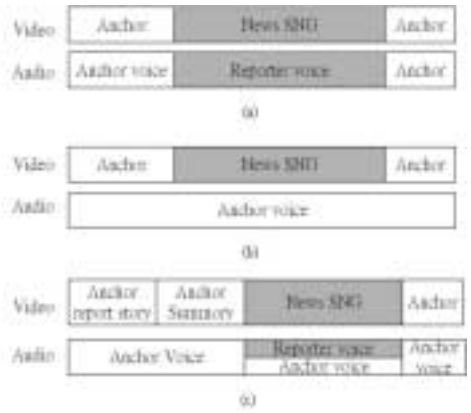

Figure 5. manifestation of anchor and non-anchor shots in video and audio domain

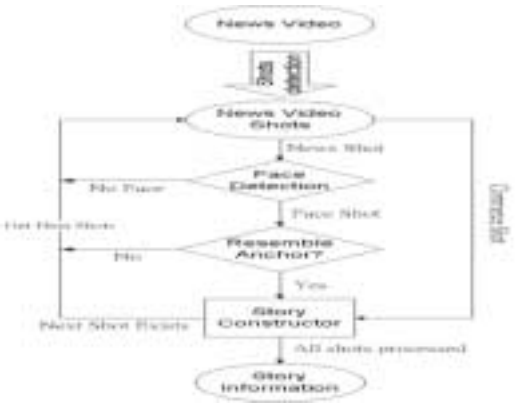

Figure 6. anchor detection and news story construction flow path

We discover that both audio and video processes can find out the boundaries between anchor and non-anchor shots under condition (a). But under condition (b), only video process can work well. And neither audio nor video process could work well under condition (c).

Hence we use video process to detect anchor shots, and the step is to detect shot change in news video first[1], then to detect if there is a face in each shot[4], and to compare the similarity between with an anchor frame and the frames

of shots that there is a face in. Afterward, we can construct the whole structure of news. It's shown in Figure 6.

4.2 Determine Anchor Shots and Compare Similarity

It's a popular research to find out specific shots from video. Anchor shots detection is especially important in news video domain[3][8]. For example, in research of [8], several anchor-shot models are trained by the spatial structure of objects in frames. But this kind of models is too confused that errors come about, for instance, someone's speech or interview will be mistaken to be anchor shots.

In our research, we bring up a simple and effective idea, that is to select the first shot which has faces and last for enough time to be the anchor shot. In general, a news program begins at an anchor shot, even if there is a summary of the day before anchor shots, it'll be destroy because of fast scene change. And we compare later face shots with this anchor shot to find out all anchor shots. The advantages of this method are easy to implement and needn't train models in advance.

At the same time, we choose color histogram to compare the similarity between two frames. Take RGB values from each pixel on frames, then the distance between between P and Q frames is:

$$\text{distance}(P,Q)_R = \sum_{k=1}^{256} \|\#(P|k)_R - \#(Q|K)_R\| \quad (4)$$

$$\text{distance}(P,Q)_G = \sum_{k=1}^{256} \|\#(P|k)_G - \#(Q|K)_G\| \quad (5)$$

$$\text{distance}(P,Q)_B = \sum_{k=1}^{256} \|\#(P|k)_B - \#(Q|K)_B\| \quad (6)$$

$$\text{distance}(P,Q) = \text{distance}(P,Q)_R + \text{distance}(P,Q)_G + \text{distance}(P,Q)_B \quad (7)$$

where $\#(P|k)_R$ means the number of pixels whose value of R is k in frame P.

The advantage of this method, except fast computation, is that even if anchor position is changed, it could still work well.

4.3 Construct News Story Structure

Up to now, we have segmented a news program into three kind of shots, anchor, non-anchor, and commercials. The structure of news stories is usually like as Figure 7. Our method of news story construction is:

1. Find out two adjacent anchor shots in order. If there is no any non-anchor shot between these two anchor shots, merge these into one(Figure 7(b) B1,B2).

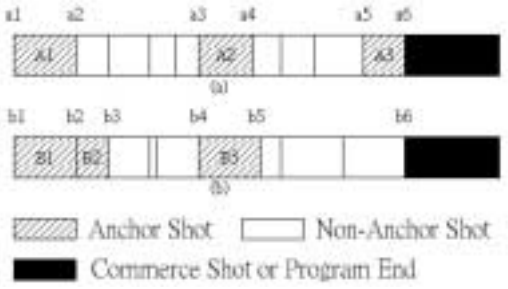

Figure 7. Structure of News Stories

2. Store the three time stamps, that is the start and end of first anchor shot and the start of second one, into database, and these stamps mean a news story. Go on constructing news stories until a commercial or the end of program.

3. When meeting a commercial,

   a. if the commercial is adjacent to the previous anchor shot, ex. Figure 7(a), reset data and continue to construct after the commercial.

   b. if there are non-anchor shots between them, construct these two units into a news story, ex. Figure 7(b), and continue to construct after the commercial.

4. When meeting the end of program, destroy the last story, because this story is a goodbye shot and meaningless.

After previous process, each story is formed by three time stamps. For example the first story in Figure 7(a) is formed by A1 and shots between A1 and A2, then the three time stamps are a1, a2, and a3. The a1 represents the start of the story, a2 is the boundary between anchor and non-anchor shots, and a3 means the end of the story. These three time swamp will be store into the columns, start_time, medium_time, and end_time, in Story_table in database.

5 Results

After news story construction, we carry out summarizing and keywords extraction for each story[1][2]. It means the background of NVBS has been finished. Then when users send request in date and channel to query, interface will return the correct table of content, including keywords and keyframes. For example, when users query CTV news on 2000,6,8, the keywords of each story on that day are put on the left of screen in order, and the keyframes are also put on the right. It is shown in Figure 5. By the way, the commercials are put on the left, too.

After clicking the keyframe or keyword of some story, more keyframes will be shown up. Users could decide if

Figure 8. table of content of CTV news on 2000,6,8

Figure 9. return of selecting any story

they want to browse this story and how to browse by these keyframes. There are four kinds of ways to browse a story: anchor shot, detailed reporting(non-anchor shots), whole of the story, and the summary. And the hyper-link is located above keyframes. It is shown in Figure 5.

The result of audio process for separating news from commercials is not bad because the special music is regular even if there is still some difference between title and end music. At the same time, the number of clusters of training model and threshold effect much, too, and we let k be 256 by experiment.

On the other hand, the result of video process for anchor detection is also very good. Table 1 shows the result.

6 Conclusions

News video must be the most important data on network in the future, but news video browsing webs are very few now. Hence we present in this paper the concept of news video browsing system and automatic construction of content hierarchy of news, including separating news from commercials, anchor detection, and story construction. After integrating with NVBS, users could decide if they are interested in this story and how to browse it by keyframes and keywords. In order that users could decide how much information he'll get from this story more freely, there are total four kinds of ways to browse it. Moreover, users could query news by date or keywords.

We plan to class all news stories automatically in the future. And we will develop a tool to let specific users modify the information of stories easily on interface.

References

[1] Chih-Chuan Chiang, *Stduy of Video Segmentation and Summary in News Video*, Master Thesis, Computer Science and Information Engineering Dept., National Chiao-Tung University, June 2000.

[2] Yu-Hang Tseng, *News Video Content Search System*, Master Thesis, Computer Science and Information Engineering Dept., National Chiao-Tung University, June 2000.

[3] Shih-Fu Chang, Qian Huang, Thomas Huang, Atul Puri, and Behzad Shahraray, "Multimedia Search and Retrieval", *Advances in Multimedia: Systems, Standards, and Networks*, 1999.

[4] Hsin-Chia Fu, P.S. Lai, R.S. Lou, H.-T. Pao, "Face Detection and Eye Localization by Neural Network Based Color Segmentation", in the *proceedings of IEEE*

Table 1. Result of Story Construction

	Program 1	Program 2	Program 3	Program 4	Program 5	Total
# of Story	23	21	23	22	22	110
# of Miss	1	0	1	3	3	8
# of Error	0	1	4	1	0	6
# of found Stories	22	21	26	20	19	108

INTERNATIONAL WORKSHOP ON NEURAL NETWORKS FOR SIGNAL PROCESSING, 11-13 December 2000, Sydney Australia.

[5] Kenichi Minami, Akihito Akutsu, Hiroshi Hamada, and Yoshinobu Tonomura, "Video Handling with Music and Speech Detection", *IEEE Multimedia*, vol.5, pp.17-25, July-Sept., 1998.

[6] P.C. Woodland, *Hidden Markov Model Toolkit V1.4 User Manual*, Speech Group, Cambridge University Engineer Department, 1992.

[7] T. Hain, S.E. Johnson, A. Tuerk, P.C. Woodland, and S.J. Young, "Segment Generation and Clustering in the HTK Broadcast News Transcription System", Speech, Vision and Robotics Group, Cambridge University Engineering Department.

[8] Hong Jiang Zhang, Shuang Yeo Tan, Stephen W. Smoliar, Gong Yihong, "Automatic parsing and indexing of news video", *Multimedia Systems*, pp. 256-266, 1995.

[9] Advanced Streaming Format(ASF) Specification, version 1.0, Microsoft Corporation, Feb 26 1998.