EXTRACTION OF DTV CLOSED CAPTION STREAM AND GENERATION OF VIDEO CAPTION FILE

Jung-Youn Kim and Jeho Nam

School of Mobile Communication & Digital Broadcasting Engineering
University of Science and Technology (UST)
Daejeon, Korea
Broadcasting & Telecommunications Media Research Department
Electronics and Telecommunications Research Institute (ETRI)
Daejeon, Korea
E-mail: goodlife@ust.ac.kr, namjeho@ust.ac.kr

ABSTRACT

This paper presents a scheme that generates a caption file by extracting a Closed Caption stream from DTV signal. Note that Closed-Captioning service helps to bridge "digital divide" through extending broadcasting accessibility of a neglected class such as hearing-impaired person and foreigner. In Korea, DTV Closed Captioning standard was developed in June 2007, and Closed Captioning service should be supported by an enforcing law in all broadcasting services in 2008. In this paper, we describe the method of extracting a caption data from MPEG-2 Transport Stream of ATSC-based digital TV signal and generating a caption file (SAMI and SRT) using the extracted caption data and time information. Experimental results verify the feasibility of a generated caption file using a PC-based media player which is widely used in multimedia service.

Keywords: DTVCC, Closed-Captioning, SAMI, SRT.

1. INTRODUCTION

Digital TV Closed-Captioning (DTVCC) is the service that shows the characters of broadcasting script. Closed-Captioning service aims to bridge "digital divide" through extending broadcasting accessibility of a neglected class such as hearing-impaired person and foreigner. In Korea, DTVCC standard was developed in June 2007 [1, 2]. It is now regulated for all broadcasting services in Korea to support DTVCC by enforcing law. Meanwhile, closed caption data is multiplexed in MPEG-2 Transport Stream (TS) [3]. To display the closed caption, specialized caption extractor and player are required.

In this paper, we show how to extract a caption data from a recorded file of ATSC-based digital TV signal and then propose a novel scheme of generating caption file that can be playbacked with a PC-based media player used in various multimedia services. The format of caption file used in the PC environment are Synchronized Accessible Media Interchange (SAMI) [4] and SubRip caption file (SRT) [5], which are widely used in the world as dominant format types for video caption file.

The rest of the paper is organized as follows. In Section 2, we explain the method of extracting closed caption data from DTV signal, and present the proposed method of generating caption file in Section 3. In Section 4, we show the experimental results by implementing both caption extractor and SAMI/SRT file generator. Finally, Section 5 concludes this paper.

This work was partly supported by the IT R&D program of MKE/IITA [2007-S-003-02 Development of Protection Technology for Terrestrial DTV Program]

2. EXTRACTION OF DIGITAL TV CLOSED CAPTION

In this section, we describe the process of extracting DTV closed caption data from MPEG-2 Transport Stream [6] in ATSC-based digital TV signal [7]. Extraction and analysis of closed caption are performed on the basis of Program and System Information Protocol (PSIP) standard [8] and Korea/international standard for DTVCC [1, 2]. Fig. 1 shows a block diagram of proposed DTV closed caption extraction and caption file generator.

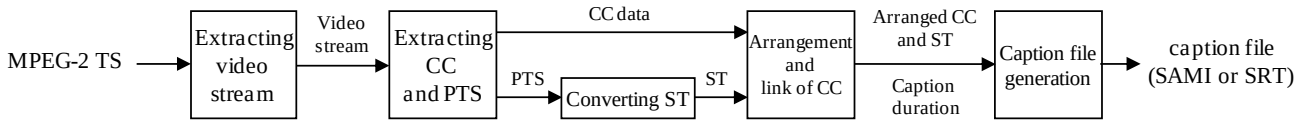
2.1 Analysis of Caption Service Descriptor

In order to extract closed caption, we begin by analyzing a Caption Service Descriptor (CSD) that contains types and attributes of caption located in Program Map Table (PMT) or Event Information Table (EIT) of PSIP [8].

Table 1 shows the syntax of CSD. The language is a 3 byte code referring language of caption, and each code of language is defined in ISO 639.2/B [9] (e.g., Korea is "kor"). Especially in Korea, there is a 1 bit field, korean_code, which specifies Hangul code with either unicode ("1") or KSC-5601 ("0"). After analysis of the other fields of CSD, received closed caption data should be interpreted based on information of CSD.

2.2 Extraction of MPEG-2 TS Video Stream

As defined in MPEG-2 Systems, MPEG-2 TS consists of 188 byte packet [3]. The type of MPEG-2 TS packet (e.g., video, audio, PSIP, etc.) can be interpreted by Packet Identifier (PID). To obtain closed caption data, we need to



* CC: Closed Caption , ST: Sync Time

Fig. 1. Block diagram of proposed DTV closed caption extraction and caption file generator.

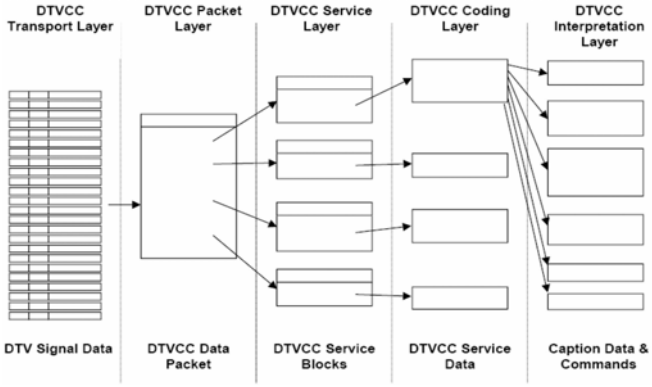


Fig. 2. DTVCC protocol model [2].

exploit a video stream because DTV closed caption data is included in the Picture User Data field of video stream. Extraction of video stream is performed by analyzing Program Association Table (PAT) and Program Map Table (PMT). Readers interested in detailed process of extraction are encouraged to refer to MPEG-2 Systems standard [6].

2.3 Extraction of Caption Data

Note that extracted video stream is composed of packetized element stream (PES), and closed caption data is located in cc_data() field within the Picture User Data of PES. Table 2 shows the syntax of cc_data(). Among the fields of cc_data(), cc_data_1 and cc_data_2 are first and second bytes of closed caption data. The value of cc_count specifies the number of cc_data_1 and cc_data_2.

Table 1. Syntax of caption service descriptor.

Syntax	No. of Bits	Format
caption_service_descriptor() {		
...		
for(i=0;i<number_of_services;i++) {		
language	8*3	uimsbf
...		
korean_code	1	bslbf
...		
}		
}		

Table 2. Syntax of cc_data().

Syntax	No. of Bits	Format
cc_data() {		
...		
for(i=0 ; i<cc_count ; i++) {		
...		
cc_data_1	8	bslbf
cc_data_2	8	bslbf
}		
}		

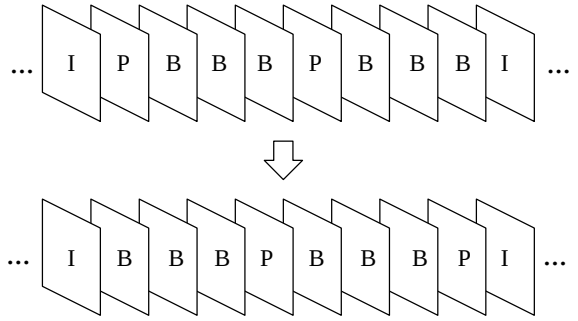


Fig. 3. Arrangement of video frames in order of PTS.

Extracted closed caption data through above process is DTVCC Packet Layer data. As shown in Fig. 2, the final closed caption data is obtained by sequentially analyzing Service Layer, Coding Layer and Interpretation Layer in the DTVCC protocol model [1, 2].

3. CAPTION FILE GENERATION

Recall that closed caption data obtained by a process described in Section 2 is converted to a caption file (i.e., SAMI or SRT). In general, caption file has sync time or caption duration corresponding to a video scene and caption data. We provide a method of how to obtain sync time, caption duration, and arranged caption data as follows.

3.1 Sync Time

In SAMI file format, there is milliseconds unit time information called sync time. Since closed caption data is in the PES of video stream, Presentation Time Stamp (*PTS*) of PES is available as the sync time information. *PTS* is 33 bit field located in the header of PES which specifies the presentation time of PES. Note that the unit of *PTS* is system clock frequency, so we convert *PTS* to sync time by

$$Sync\ time = (PTS/90) - (PTS_{start}/90). \quad (1)$$

In order to convert *PTS* to second unit, it is required to divide *PTS* by 90kHz. However, since sync time needs millisecond unit value, *PTS* should be divided by 90Hz. PTS_{start} means the *PTS* of the first PES of video stream. Note that video stream is transmitted by not the order of *PTS* but the order of decoding time of each frame. Therefore, it needs to arrange video frames through *PTS* as illustrated in Fig. 3.

3.2 Arrangement and Link of Caption

To arrange completed words and sentences properly, it

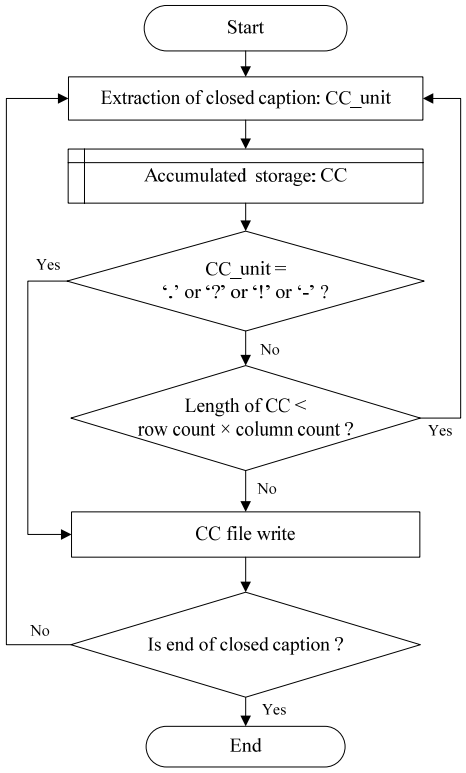


Fig. 4. The flow chart of linking closed caption.

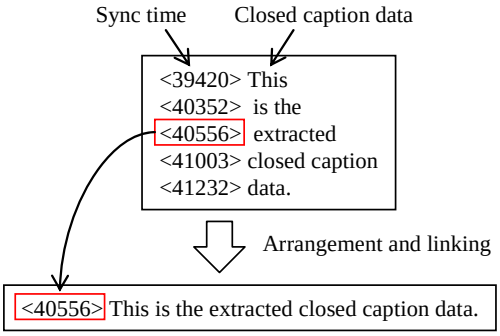


Fig. 5. Selection of final sync time.

needs the operation of liking caption data. As a criterion of the number of row and column displayed to window, DefineWindow that is one of the Command descriptor of Interpretation Layer is available [2]. The row count and column count specify the numbers of row and column displayed respectively. The row lock and column lock indicate whether the values of row count and column count are either fixed number the window contains or not respectively. When set to YES ("1") in row lock or column lock, caption has to show corresponding to the value of row count and column count. However, when set to NO ("0") the values of row count and column count are not absolute value to show the closed caption.

Proposed method considers only the case that row lock and column lock set to NO. Thus, we use row count and column count as maximum length of closed caption data corresponding to each sync time. Fig. 4 shows the flow chart of linking closed caption. At first, closed caption data is extracted from each PES. Then the closed caption data is stored to a temporal buffer. Next, accumulated closed caption data are analyzed to write a file. When accumulated closed caption data meets the conditions as described in the

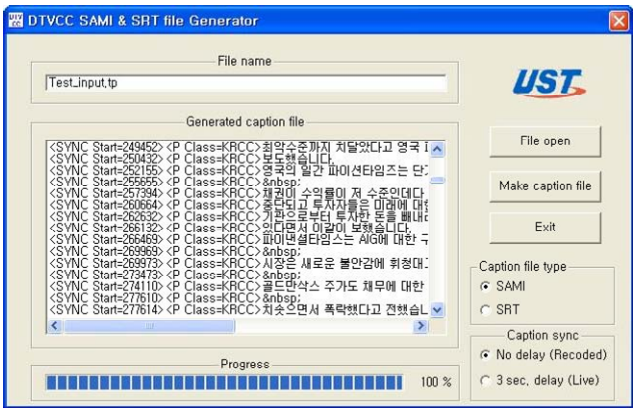


Fig. 6. DTVCC SAMI & SRT file generator.

conditional statements of Fig. 4, accumulated closed caption data are outputted to the caption file.

Getting each closed caption data together by above criterion, we need to select one sync time among some sync times corresponding to each closed caption data. We take a median sync time for final sync time as illustrated in Fig. 5. The rectangle above in Fig. 5 contains sync time and closed caption data which extracted from each PES, and the rectangle below in Fig. 5 shows an example of a finally selected sync time and a concatenated caption data.

3.3 Caption Duration

To make a SRT file format, the information of caption duration is required. Caption duration for SRT file consists of caption start time and end time. Both start time and end time consist of hour, minute, and second down to three decimal places (i.e., millisecond unit). While start time is easily obtained from sync time in Section 3.1, we propose the method of computing end time as follows:

$$End\_time(i) = (NW \times 60) / \alpha + Start\_time(i),$$
$$End\_time(i) < Start\_time(i+1). \quad (2)$$

NW is the number of words in caption data, and α is a kind of scaling factor which means the number of words spoken per one minute. The effective value of α can be selected variably by user experiments. The end time in which index number is i must be lower than the next start time of caption in which index number is $i+1$.

Obtained caption duration should be used to generate a SRT file.

4. IMPLEMENTATION OF SYSTEM

In order to verify the feasibility of our proposed technique, we implement the caption file generator of DTV closed caption data using Microsoft Visual C++ 6.0 MFC, which is illustrated in Fig. 6. Input stream is DTV program that is recorded through HDTV card. Since digital broadcasting stream is transmitted by MPEG-2 TS, the format of input stream is MPEG-2 TS. The "Caption sync" shown in Fig. 6 sets up the delay of caption. In Korean stenographic system, there is 2~3 seconds delay on live broadcasting (e.g., News, TV debate, etc.). Therefore,

<SYNC Start=245248><P Class=KRCC>
앵커: 전세계 공황상태가 2차세계대전 이후
<SYNC Start=248748><P Class=KRCC>
<SYNC Start=249452><P Class=KRCC>
최악수준까지 치달았다고 영국 파이낸셜타임스가
<SYNC Start=250432><P Class=KRCC>
보도했습니다.
<SYNC Start=252155><P Class=KRCC>
영국의 일간 파이션타임즈는 단기 상환 미국 재무부
<SYNC Start=255655><P Class=KRCC>
<SYNC Start=257394><P Class=KRCC>
채권이 수익률이 저 수준인데다 은행간 대출도
<SYNC Start=260664><P Class=KRCC>
중단되고 투자자들은 미래에 대한 불안감으로
<SYNC Start=262632><P Class=KRCC>
기관으로부터 투자한 돈을 빼내려고 아우성치고
<SYNC Start=266132><P Class=KRCC>
있다면서 이같이 보했습니다.
<SYNC Start=266469><P Class=KRCC>
파이낸셜타임스는 AIG에 대한 구제금융조치에도
<SYNC Start=269969><P Class=KRCC>
<SYNC Start=269973><P Class=KRCC>
시장은 새로운 불안감에 휘청대고 있고 모건스텐리

(a)

26
00:04:05,248 --> 00:04:07,748
앵커: 전세계 공황상태가 2차세계대전 이후

27
00:04:09,452 --> 00:04:11,452
최악수준까지 치달았다고 영국 파이낸셜타임스가

28
00:04:10,432 --> 00:04:10,932
보도했습니다.

29
00:04:12,155 --> 00:04:15,655
영국의 일간 파이션타임즈는 단기 상환 미국 재무부

30
00:04:17,394 --> 00:04:20,394
채권이 수익률이 저 수준인데다 은행간 대출도

31
00:04:20,664 --> 00:04:23,164
중단되고 투자자들은 미래에 대한 불안감으로

(b)

26
00:04:05,248 --> 00:04:07,748
アンカー: 全世界 恐慌状態が 2次世界大戦 以後

27
00:04:09,452 --> 00:04:11,452
最悪水準まで 駆け上がったと 英国 Financial Timesが

28
00:04:10,432 --> 00:04:10,932
報道しました..

29
00:04:12,155 --> 00:04:15,655
英国の 日刊 Financial Timesは 短期 償還 米国 財務部

30
00:04:17,394 --> 00:04:20,394
債権が 収益率が あの 水準なのに加え 銀行間 貸出も

31
00:04:20,664 --> 00:04:23,164
中断になって 投資者らは 未来に対する 不安感で(に)

(c)

Fig. 7. Generated SAMI and SRT caption file: (a) SAMI file (Korean), (b) SRT file (Korean), and (c) SRT file (Japanese).

Fig. 8. Test News program applied caption file.

when we generate a caption file from recorded file of live broadcasting, we should apply 3 seconds delay to caption file. In case of a filmed TV broadcast, no delay is needed.

We tested Korea News program as input data. We generated both SAMI and SRT files by extracting closed caption data and by computing sync time and caption duration. We set α to 120 for SRT file. The other parameters set to default values such as text color, size, etc.

Fig. 7 shows the results of generating caption file and application. Fig. 7(a) and Fig. 7(b) show the results of generating SAMI and SRT files respectively. Moreover, Fig. 7(c) is an example of the application service using a caption file: Generated caption data can be translated either manually or automatically to other languages. The translation of Korean-to-Japanese is provided in Fig. 7(c). We use an automatic translation tool, TransCAT [10]. Fig. 8 shows a part of TV news program with closed caption data playbacked in multimedia player of the PC environment. Experimental results show that generated caption file performs well in general PC multimedia players such as the Windows Media Player.

5. CONCLUSIONS

This paper proposed a technique of extracting closed caption data from ATSC-based digital TV signal and of generating a caption file that is usable in PC environment. The formats of generated caption file are SAMI and SRT, which are widely used in multimedia service. We confirmed the feasibility of generated caption file works with general multimedia player. Note that extracted caption data is highly useful in a variety of multimedia application. For example, caption data can be translated either manually or automatically to foreign languages and provided as a value-added content, and used for video indexing/retrieval. We plan to study on various applications based on exploitation of DTV closed caption data afterward.

6. REFERENCES

[1] "Standard for Terrestrial Digital TV Closed Captioning," TTA, TTAS. KO-07.0050, Jun. 2007.

[2] "Digital Television (DTV) Closed Captioning," EIA-708-B, Dec. 1999.

[3] "Digital Television Standard, Part 2 - MPEG-2 Video System Characteristics," Doc. A/53, Part 4:2007, Advanced Television Systems Committee, Jan. 2007.

[4] "Understanding SAMI 1.0," Microsoft Developer Network (MSDN), http://msdn2.microsoft.com/en-us/library/ms971327.aspx, Feb. 2003.

[5] "The Brain's Web," http://membres.lycos.fr/subrip.

[6] "Information technology -- Generic coding of moving pictures and associated audio information: Systems," ISO/IEC International Standard 13818-1, 2000.

[7] "ATSC Digital Television Standard," Advanced Television Systems Committee, Doc. A/53E, Dec. 2005.

[8] "Program and system information protocol for terrestrial broadcast and cable (Revision C) with amendment No. 1," Advanced Television Systems Committee, Doc. A/65C, Jan. 2006.

[9] "Code for the Representation of Names of Languages — Part 2: alpha-3 code," ISO 639.2, as maintained by the ISO 639/Joint Advisory Committee (ISO 639/JAC), http://www.loc.gov/standards/iso639 -2 /iso639jac.html.

[10] TransCAT KJJK, http://www.dicosystem.com/.