

특집논문 (Special Paper)

방송공학회논문지 제21권 제3호, 2016년 5월 (JBE Vol. 21, No. 3, May 2016)

<http://dx.doi.org/10.5909/JBE.2016.21.3.307>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

뉴스 비디오 시퀀스에서 텍스트 시작 프레임 검출 방법의 비교

이상희^{a)}, 안정일^{b)}, 조강현^{a)†}

Comparison of Text Beginning Frame Detection Methods in News Video Sequences

Sanghee Lee^{a)}, Jungil Ahn^{b)}, and Kanghyun Jo^{a)†}

요 약

비디오 프레임 내의 오버레이 텍스트는 음성과 시각적 내용에 부가적인 정보를 제공한다. 특히, 뉴스 비디오에서 이 텍스트는 비디오 영상 내용을 압축적이고 직접적인 설명을 한다. 그러므로 뉴스 비디오 색인 시스템을 만드는데 있어서 가장 신뢰할 수 있는 실마리이다. 텔레비전 뉴스 프로그램의 색인 시스템을 만들기 위해서는 텍스트를 검출하고 인식하는 것이 중요하다. 이 논문은 뉴스 비디오에서 오버레이 텍스트를 검출하고 인식하는데 도움이 되는 오버레이 텍스트 시작 프레임 식별을 제안한다. 비디오 시퀀스의 모든 프레임이 오버레이 텍스트를 포함하는 것이 아니기 때문에, 모든 프레임에서 오버레이 텍스트의 추출은 불필요하고 시간 낭비이다. 그러므로 오버레이 텍스트를 포함하고 있는 프레임에만 초점을 맞추으로써 오버레이 텍스트 검출의 정확도를 개선할 수 있다. 텍스트 시작 프레임 식별 방법에 대한 비교 실험을 뉴스 비디오에 대해서 실시하고, 적절한 처리 방법을 제안한다.

Abstract

Overlay texts are artificially superimposed on the broadcasting videos by human producers. These texts provide additional information to the audiovisual content. Especially, the overlay texts in news video contain concise and direct description of the content. Therefore, it is most reliable clue for constructing a news video indexing system. To make this indexing system in the TV news program, it is important to detect and recognize the texts. This paper proposes the identification of the overlay text beginning frame to help the detection and recognition of the overlay text in news videos. Since all frames in the video sequences do not contain the overlay texts, the overlay text extraction from every frame is unnecessary and time-wasting. Therefore, to focus on only the frame containing the overlay text can be enhanced the accuracy of the overlay text detection. The comparative experiments of the text beginning frame identification methods were carried out with respect to Korean television news videos. Then the appropriate processing method is proposed.

Keyword : overlay text, beginning frame, edge detector, news video, indexing

a) 울산대학교 전기공학부(School of Electrical Engineering, University of Ulsan)

b) ㈜울산방송(Ulsan Broadcasting Corporation)

† Corresponding Author : 조강현(Kanghyun Jo)

E-mail: acejo@ulsan.ac.kr

Tel: +82-52-259-2208

ORCID: <http://orcid.org/0000-0001-8317-6092>

※ 이 논문의 연구결과 중 일부는 "IWAIT 2016"에서 발표한 바 있음.

· Manuscript received March 15, 2016; Revised April 25, 2016; Accepted May 4, 2016.

I . INTRODUCTION

Nowadays, the rapid growth of video data has motivated many researchers to design and develop an efficient content-based browsing and retrieving system. In response to such needs, various video content analysis schemes using one or a combination of image, audio, and textual information in the videos have been proposed to parse, index, or abstract massive amounts of data. Among these information sources, the texts present in the video frames can provide important supplemental information for indexing and retrieval. For example, with the help of the extracted text related to the news, the news videos can be segmented and catalogued more accurately in the sense of semantics. Therefore, the extraction of video text information has a very important significance for the further semantic understanding^[1,2].

In general, the text in video sequences can be divided into the overlay text and the scene text. The scene text naturally exists in the image being recorded in native environment. This text is found in street signs, text on trucks, and the writing on shirts in natural scenes. The appearance of the scene text is occasional, and the difference among different the scene texts is very big. On the other hand, the overlay text, is called graphics text or caption in other papers, is graphically generated and artificially overlaid on the image by human at the time of editing. This text is used to supplement visual or audio content such as the subtitles in news video, sports scores. The overlay text has three characteristics: the first, in foreground, the second, the color is independent with the background, and the third, the text distributes in a rule region vertically or hori-

zontally in general. Although these two type texts often contain important information about the content of the video, the overlay text contains more concise and direct description. Therefore, the detection and recognition of the overlay text is an essential issue on the automated content analysis systems^[2-7].

Many research projects have engaged in the overlay text detection and extraction in video sequences. As shown in Fig. 1, the extraction system of the overlay text information from video sequences mainly consists of four main parts: text detection, localization, segmentation, and recognition. Text detection is to find the text regions if there is text in a video frame. Text localization is to group text regions into text lines and generate a set of tight bounding boxes around all text lines. The text segmentation or tracking is to determine the temporal and spatial locations. Text recognition is segmented for text regions by binary image and performed the OCR (Optical Character Recognition) system^[2,4,5].

Although today's state-of-the-art OCR methods are accurate, most of all, the extraction system requires a good detection of the text regions in image among the system consisting parts. The text detection step must find a maximum amount of text, but also find the exact coordinates of the boxes that contain text. However, low resolution of the imagery, the richness of the background, and compression artifacts limit the detection accuracy that can be achieved in practice using existing text detection algorithms. Without an accurate surrounding box, the quality of text recognition is degraded, leading to poor performances. Current video text detection approaches can be classified into two categories. One is detecting text regions individual frames



그림 1. 오버레이 텍스트 추출 시스템의 구조
 Fig. 1. Structure of the overlay text extraction system

independently. This category can be divided into the connected component based methods, texture analysis based methods, and gradient edge based methods. The other is utilizing the temporality of the video sequences, and it is based on the fact that the overlay texts generally last at the same position for a few seconds^[3-10]. The proposed paper aims to achieve a good accuracy of video text detection by temporal analysis. Especially, this paper focuses on detecting the overlay text in news video sequences.

The overlay text in news videos provides more meaningful information of the content than any other type of videos. For instance, the text annotates the names of people and places, or describes objects and the current issue. Since the broadcasting videos are produced by professional, the overlay text of the TV news program uses the rule-based characteristics. The common properties in most news video sequences are summarized as follows: The overlay text position is fixed, generally in the range of 1/3 from the bottom of the frame. The background of the text is usually opaque or translucent matte, and in most case, the color of the background matte is eye-catching, such as white, blue, yellow, and so on. Colors of text character are often very distinguishable from the background color. The size and font of the overlay texts in the same news video generally remain unchanged for a long term^[11].

In many of the previous researches for the overlay text detection in the video sequences, their approaches aimed to extract the text from every frame of the video sequences.

However, these approaches are not adequate to news videos. As shown in Fig. 2, the overlay texts in video sequences are the appearances and disappearances of the overlay text occur suddenly or slowly. Since the overlay texts appearing or disappearing in news videos are not changes faster than scene content in a shot, the overlay text extraction from every frame is unnecessary and time-wasting. By precisely finding the critical frame where each overlay text appears or disappears, the overlay text detection can be focused solely on the frame containing the overlay texts. Therefore, to achieve a good result of the overlay text detection in news video sequences, this paper proposes the identification of the overlay text beginning frame. The beginning frame is defined as the frame which has abrupt difference at edge density of current frame and previous frame and a little difference at the edge density of current frame and next frame. The proposed method acts as the pre-processing step for the output of a text detector in the entire video sequences.

In Ref. [7], a multi-resolution change detection algorithm is used along the time axis to detect the appearance and disappearance of multiple, concurrent lines of the text followed by recursive time-averaged projections in Y and X axes. In Ref. [8], the overlay text beginning frame is detected first. And then, according to appearance frequency of the likely text block, the overlay text candidate region in the overlay text beginning frame is defined. Since texts contain rich edge information, this paper uses the edge den-

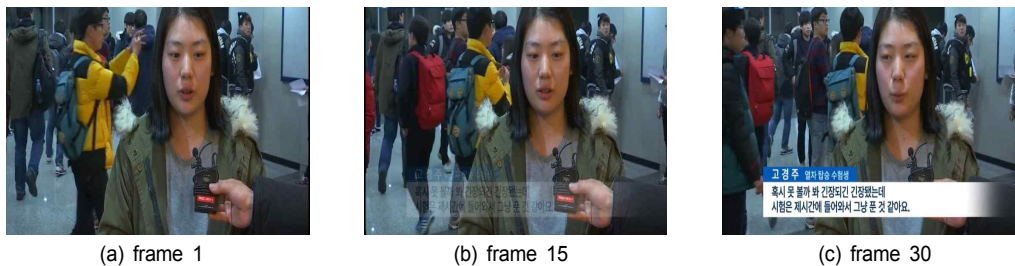


그림 2. 뉴스 비디오 시퀀스에서 오버레이 텍스트의 나타남 예
Fig. 2. Example of overlay text appearances in news video sequences

sity of the text obtained by edge detector. If a frame has a high edge density, it has the overlay text. Otherwise, a frame does not contain the overlay text. The edge based method used in this paper is the Canny edge detector. In the experiments, this method's performance of the beginning frame identification are compared with the Harris corner detector. After then, the appropriate processing method is proposed. The section III explains the two methods to identify the overlay text beginning frame in detail. Section IV shows and analyzes the results of the experiments. And the last section describes the brief conclusion and the future works.

II. RELATED WORKS

The many previous text detection and recognition methodologies have been classified commonly two categories. One is the stepwise methodology and the other is the integrated methodology. As shown in Fig. 3(a), the stepwise methodologies have separated detection and recognition modules, and use a feed-forward pipeline to detect, segment and recognize text regions. Some stepwise ap-

proaches utilize a feedback procedure from text recognition to reduce false detections. By contrast, the integrated methodologies have a goal of recognizing word where the detection and recognition procedures share information with character classification and/or use joint optimization strategies, as shown in Fig. 3(b). Some integrated approaches use a pre-processing step to localize regions of interest. The key difference lies in the fact that the latter uses recognition as a key focus^[3].

And current video text detection approaches can be classified into two categories. One category is detecting text regions in individual frames independently. The other is utilizing the temporality of the video sequences. The first category can be further divided into three kinds: connected-component-based methods, texture-analysis-based methods, and gradient-edge based methods. The last category is based on the fact that the overlay texts generally last at the same position for a few seconds^[6].

Most approaches for text detection and localization in video sequences produce an estimate of the bounding box for each text line in individual frames. Li et al.^[12] extracted wavelet texture features from image blocks and used a neural network to discriminate the text from the non-text

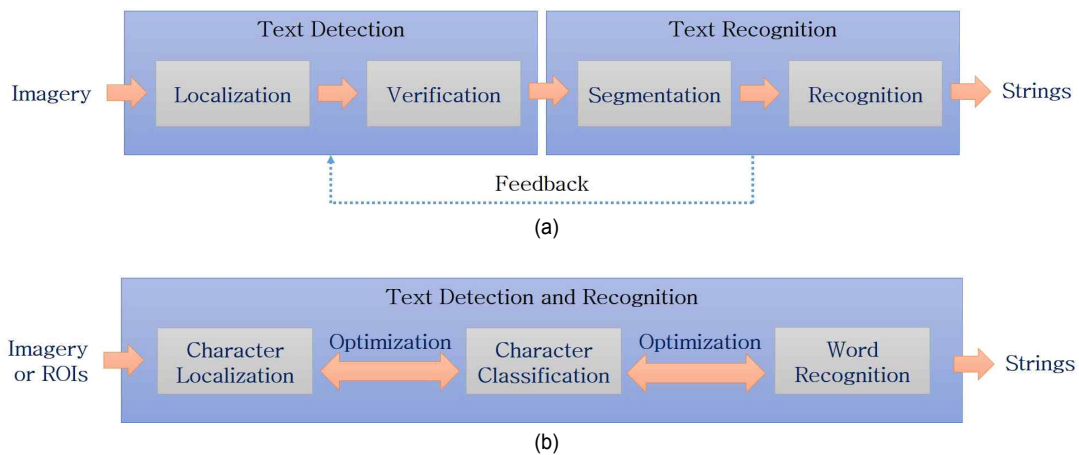


그림 3. 텍스트 검출 및 인식에 사용된 일반적인 두 가지 방법의 프레임워크^[3]. (a) 단계별 방법 (b) 통합적 방법
 Fig. 3. Frameworks of two commonly used text detection and recognition methodologies^[3]. (a) Stepwise methodology (b) Integrated methodology

blocks. They also made use of the fact that the text remains in the scene for many consecutive frames to reduce the processing time. Lienhart et al.^[13] used gradient features and a multi-layer feed-forward network as the classifier. In algorithms of Li et al.^[12], Lienhart et al.^[13], and Lyu et al.^[14], video temporal redundancy was considered in the post-processing. Tang et al.^[15] proposed a spatial-temporal approach for video caption detection. They detected the video shot boundary first, and then recognized the caption from the caption transition frames.

This paper focuses on improving the result of text detection step which is most important at the whole extraction system. Especially, this work aims to find the beginning frame in video sequences for the pre-processing of the text detection step. By applying the text detection on only the overlay text including frames, the accuracy of text detection can be improved compared to the methods without the pre-processing.

In order to detect the beginning frame, this paper uses the discriminative properties of text and its basic unit, character. Since text regions contain rich edge information, and the pre-processing has not to waste the processing time of the whole system, an edge based extractor is efficient to identify the beginning frame and reduce the processing time. The edge detection process serves to simplify the analysis of image by drastically reducing the amount of data to be processed, while at the same time preserving useful

structural information about object boundaries^[17]. And the Canny edge detector is used in our previous work [2] which detects the overlay name text line. To prove that the Canny edge detector can be properly used to detect the beginning frame, this paper has been compared with the Harris corner detector of the corner based method. Thus, this paper contributes to help the proper pre-processing method's selection for achieving good accuracy for text detection.

III. PROPOSED METHOD

By observing a large quantity of the TV news programs, as shown in Fig. 4, the appearances and disappearances of the overlay text occur suddenly or slowly in most news videos. All frames in the video sequences do not contain the overlay texts. Therefore, this paper considers the fact that the input video frames can be divided three periods; non-text period, transition period, and text period.

Based on the above characteristics, this paper discusses the identification of the overlay text beginning frame to help the detection and recognition of the overlay text in news video sequences. Since all frames do not contain the overlay text, the detection of the overlay text in every frame is time consuming. If the detection and recognition is limited only at the frames superimposed onto the graphical

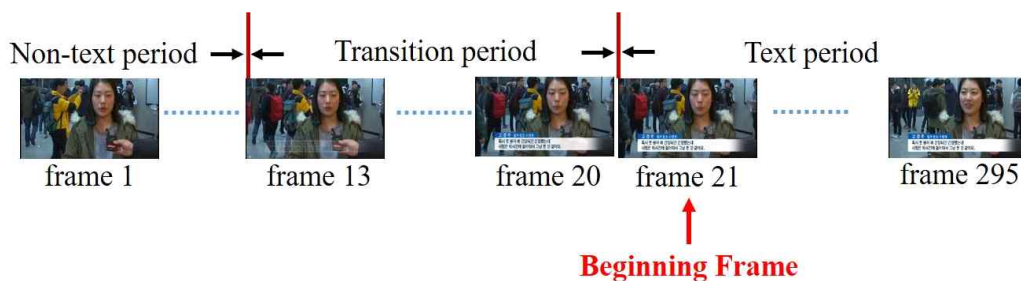


그림 4. 뉴스 비디오 시퀀스에서 오버레이 텍스트의 특성
 Fig. 4. Characteristics of overlay text in news video sequences

text, the whole processing time can be saved and the text detection accuracy can be increased. This paper uses the edge density for the identification of the overlay text beginning frame. To detect the edge in a frame, the Canny edge detector is used, and its performance of the beginning frame identification is compared with the Harris corner detector. The basic idea of the compared methods is based on the observation that the text regions typically are rich of corners and edges. And corners and edge points are nearly uniformly distributed in text areas.

1. Detection of Beginning Frame with Edge

Since texts are composed of line segments and text regions contain rich edge information, a edge based method is used to extract the text in video sequences. This paper

uses the Canny edge detector to extract the edge points.

The Canny edge detector is based on the specification of detection and localization criteria in a mathematical form. It is necessary to augment the original two criteria with a multiple response measure in order to fully capture the intuition of good detection. The detector uses adaptive thresholding with hysteresis to eliminate streaking of edge contours. The thresholds are set according to the amount of noise in the image, as determined by a noise estimation scheme. This detector made use of several operator widths to cope with varying image signal-to-noise ratios, and operator outputs were combined using a method called operators were used to predict the large operator responses. If the actual large operator outputs differ significantly from the predicted values, new edge points are marked. It is therefore possible to describe edges that occur at different

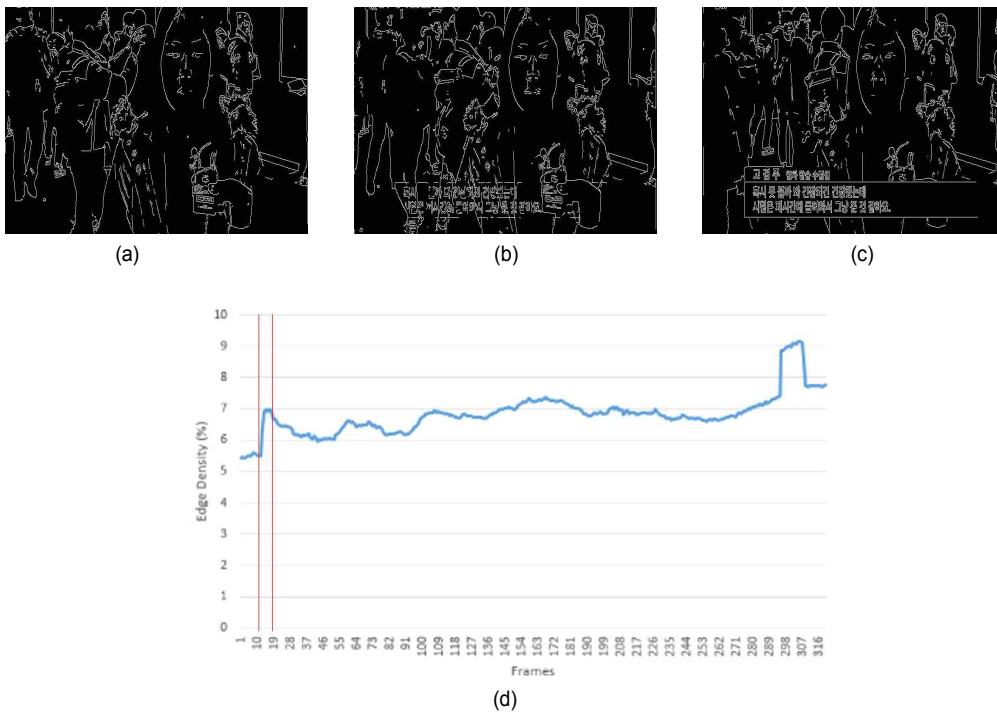


그림 5. 캐니 에지 검출기를 사용한 에지 밀도 (a) 비텍스트 구간 (b) 전환 구간 (c) 전체 비디오 시퀀스에서 에지 밀도 그래프
 Fig. 5. Edge density using Canny edge detector (a) non-text period (b) transition period (c) text period (d) plot of edge density in whole video sequences

scales, even if they are spatially coincident^[17].

Since the text presents many edges, the frame including the overlay text has significant changes in the text edge density than the frame not the overlaid text as shown in Fig. 5. The period between the vertical red lines, in other words from the frame 13 to the frame 20 like Fig. 5(d) has abruptly different from the previous frame and the edge density of the frame 21 has little different of the edge density of the frame 22. Thus, the non-text period is from the frame 1 to the frame 12, the transition period from the frame 13 to the frame 20, the text period is from the frame 21. As a result, the beginning frame becomes the frame 21. The edge density defines as the number of detected edge pixels in a frame over the total number of pixels in the

frame.

2. Detection of Beginning Frame with Corner Points

Corner points are the image features that are usually more salient and robust than edge for pattern representation. A corner can be defined as the intersection of two edges or a point where there are two dominant and different edge directions in a local neighborhood of the point. This paper uses the Harris corner detector to extract the corner points. The method is based on the local auto-correlation function of a signal, which measures the local change of the signal with patches shifted by a small amount in different directions^[18].

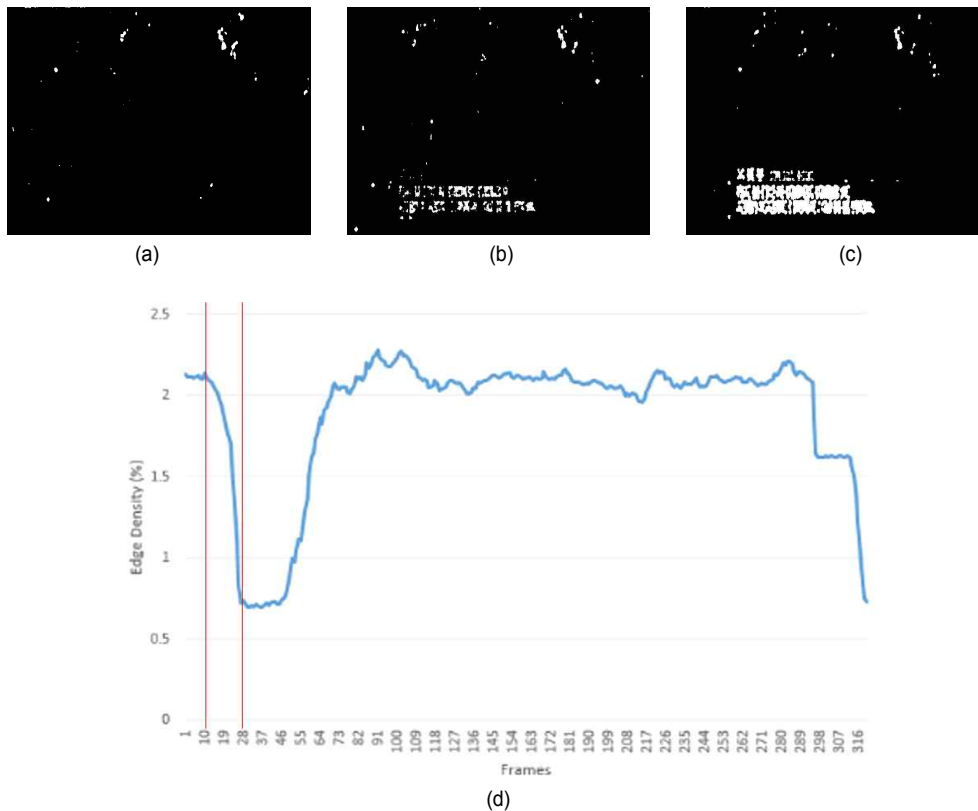


그림 6. 해리스 코너 검출기를 사용한 에지 밀도 (a) 비텍스트 구간 (b) 전환 구간 (c) 전체 비디오 시퀀스에서 에지 밀도 그래프
 Fig. 6. Edge density using Harris corner detector (a) non-text period (b) transition period (c) text period (d) plot of edge density in whole video sequences

Fig. 6 shows the result images with the detected corner points and the edge density plot. The period between red vertical lines as shown in Fig. 6(d), in other words from the frame 12 to the frame 28, has abruptly different of the edge density. Thus, the non-text period is from the frame 1 to the frame 11, the transition period is from the frame 12 to the frame 28, the text period is from the frame 29 as shown in Fig. 6(d). As a result, the beginning frame is the frame 29.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

This paper compared and analyzed two methods to identify the beginning frame. And it was tested for the effectiveness of the proposed method at our previous work [2] in two cases; i.e. using the beginning frame, and not. Since

there is no standard dataset for the proposed method, the test videos used in the experiment were captured from the video sequences in TV news program in Korea. The resolution of the videos was 720×480 . The tunable sensitivity parameter k of Harris corner response function is an empirically determined constant from 0.04 to 0.06^[18,19]. In this experiment, the parameter k set to be 0.04. And Two thresholds of Canny edge detector, i.e. T_{high} and T_{low} was decided based on our empirical studies^[2,19]. Section IV. 1 presented the results of the identification of the overlay text beginning frame using two methods as shown in Fig. 7 and table 1. The reference value of abrupt difference among the frames was decided that T_{Canny} of Canny edge detector was 0.03 and T_{Harris} of Harris corner detector was 0.003, respectively. These reference values were based on our empirical studies. Section IV. 2 showed the results that the accuracy of the overlay text detection using the beginning frame was tested at our previous work [2].

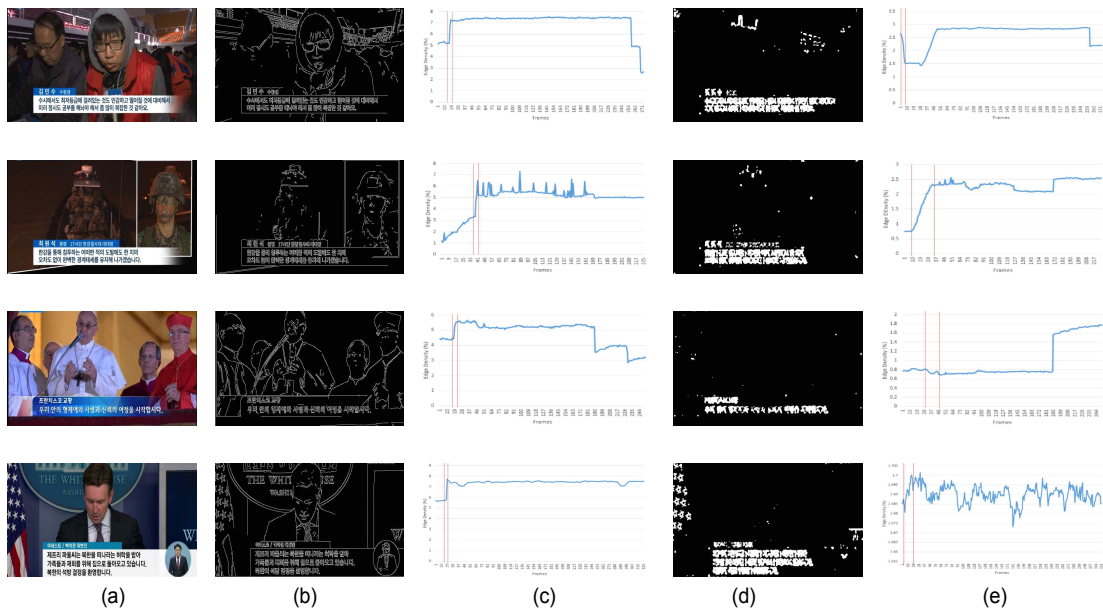


그림 7. 실험 결과 예 (a) 오버레이 텍스트 원본 이미지 (b) (a) 이미지의 캐니 에지 이미지 (c) 캐니 에지 검출기를 이용한 에지 밀도 그래프 (d) (a) 이미지의 해리스 코너 이미지 (e) 해리스 코너 검출기를 이용한 에지 밀도 그래프

Fig. 7. Examples of experimental result (a) Original image with overlay text, (b) Canny edge image of (a), (c) Plot of edge density using Canny edge detector, (d) Harris corner image of (a), (e) Plot of edge density using Harris corner detector

1. Comparison of the Two Beginning Frame Identification Methods

In Fig. 7, the first column (a) is the original image superimposed onto the overlay text in video sequences. The second column (b) is the Canny edge image of (a), the third column (c) is the edge density plot using the Canny edge detector in whole video sequences of (a). The fourth column (d) is the Harris corner image of (a), and the last column (e) is the edge density plot using the Harris corner detector in whole video sequences of (a).

As shown in Table 1, the result of the first row in Fig. 7 shows that the transition period is from the frame 16 to the frame 27 and the beginning frame becomes the frame 28 in case of using the Canny edge detector. As using the Harris corner detector, the test 1 shows that the transition period is from the frame 1 to the frame 11 and the beginning frame becomes the frame 12. The result of the Canny edge detector is similar to the that of the ground truth. On the contrary, the result of the Harris corner detector is wrong.

The result of the second row in Fig. 7 shows that the transition period is from the frame 30 to the frame 42 and the beginning frame becomes the frame 43 in case of using the Canny edge detector. As using the Harris corner detector, the test 2 shows that the transition period is from the frame 10 to the frame 33 and the beginning frame becomes the frame 34. The result of Canny edge detector is

similar to the that of the ground truth. On the contrary, the result of the Harris corner detector is wrong.

The result of the third row in Fig. 7 shows that the transition period is from the frame 14 to the frame 24 and the beginning frame becomes the frame 25 in case of using the Canny edge detector. As using the Harris corner detector, the test 3 shows that the transition period is from the frame 32 to the frame 45 and the beginning frame becomes the frame 46. The result of the Canny edge detector is similar to the that of the ground truth. On the contrary, the result of the Harris corner detector is wrong.

The result of the fourth row in Fig. 7 shows that the transition period is from the frame 18 to the frame 22 and the beginning frame becomes the frame 23 in case of using the Canny edge detector. As using the Harris corner detector, the test 4 shows that the transition period is from the frame 2 to the frame 19 and the beginning frame becomes the frame 20. The result of the Canny edge detector is similar to the that of the ground truth. On the contrary, the result of the Harris corner detector is wrong.

The Canny edge detector method relatively well detects the beginning frame than the Harris corner detector. Since the value of edge density using the Harris corner is smaller than that of using the Canny edge detector, the result shows that the Canny edge detector is more sensitivity than the Harris corner detector, and its value deviation is bigger. As a result, the decision performance of the Harris corner detector is low and it cannot be used, whereas the beginning

표 1. 캐니 에지 검출기와 해리스 코너 검출기의 비교
 Table 1. Comparison of Canny edge detector with Harris corner detector

	Ground Truth		Canny Edge Detector		Harris Corner Detector	
	Transition frame	Beginning frame	Transition frame	Beginning frame	Transition frame	Beginning frame
Test 1	16~24	25	16~27	28	1~11	12
Test 2	38~47	48	30~42	43	10~33	34
Test 3	18~25	26	14~24	25	32~45	46
Test 4	18~24	25	18~22	23	2~19	20

frame decision performance of the Canny edge detector is good. Thus, the Canny edge detector is proper as the pre-processing for the text detection step.

2. Experiments of the Beginning Frame Effectiveness

To prove that the beginning frame identification is effective and useful, this paper was experimented in our pre-

vious work [2] which detects the overlay name text to make the automatic person indexing of interview video in TV news program.

To detect the overlay name text in news video sequences, Fig. 8 shows the comparative experiment results in case of using the beginning frame, and not. As shown in Fig. 8(b), the overlay name text was properly detected by using the identification of the beginning frame based on the Canny edge detector. In contrast to, the experiments which are not

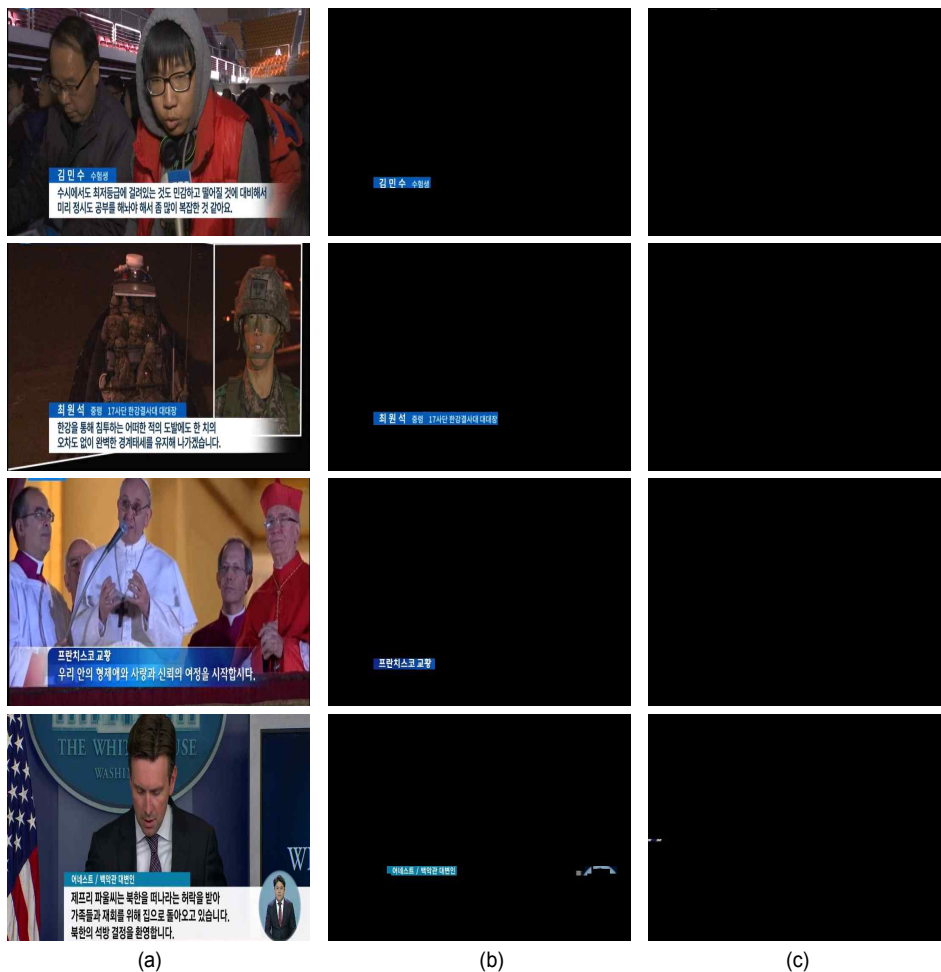


그림 8. 시작 프레임을 사용한 경우와 사용하지 않은 경우의 비교 실험 결과 예 (a) 시작 프레임 이미지 (b) 시작 프레임을 사용한 오버레이 텍스트 검출 결과 (c) 시작 프레임을 사용하지 않은 텍스트 검출 결과

Fig. 8. Examples of comparative experiment result using the beginning frame, or not (a) beginning frame image (b) the result of detected name overlay text using the beginning frame (c) the result not using the beginning frame

using the beginning frame identification fail to detect the overlay name text, like Fig. 8(c). Therefore, the result of the beginning frame identification method helps to accurately detect the overlay text.

V. CONCLUSION

This paper proposes the identification of the overlay text beginning frame to help the detection and recognition system of the overlay text in the news interview video sequences. To decide the beginning frame, the edge density of text is used by the Canny edge detector. And this paper proves that the beginning frame decision performance of the Canny edge detector is better than that of the Harris corner detector. The effectiveness and usage of the proposed method proves through the experiments of our previous work [2]. Therefore, the proposed method helps to save the whole processing time and to enhance the detection step result of the overlay text.

For readability in a complex scene, the overlay text generally is superimposed on the opaque or translucent background matte. This transparency ratio in the overlay text region can be used to more accurately detect the beginning frame. Since the transition period has continuously changed the transparency ratio and not changed after the beginning frame. Thus, for the automatic news video indexing, this property can be used to help the detection and recognition of the overlay text in news video sequences. And this method will be worked in the future.

참 고 문 헌 (References)

- [1] Xian-Sheng Hua, Liu Wenjin, Hong-Jiang Zhang, An Automatic Performance Evaluation Protocol for Video Text Detection Algorithms, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, No. 4, pp. 498-507, April 2004.
- [2] Sanghee Lee, Jungil Ahn, Kanghyun Jo, Automatic Name Line Detection for Person Indexing Based on Overlay Text, *Journal of Multimedia and Information System*, Vol. 2, No. 1, pp. 163-170, March 2015.
- [3] Qixiang Ye, DAvid Doermann, Text Detection and Recognition in Imagery: A Survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 37, No. 7, pp. 1480-1500, July 2015
- [4] Zhujun Wang, Xiaoyu Wu, Lei Yang, Ying Zhang, A Survey on Video Caption Extraction Technology, *The 4th International Conference on Multimedia Information Networking and Security*, pp. 713-716, 2012.
- [5] Jing Zhang, Rangachar Kasturi, Extraction of Text Objects in Video Documents: Recent Progress, *The 8th IAPR Workshop on Document Analysis Systems*, pp. 5-17, 2008.
- [6] Jiamin Xu, Palaiahakote Shivakumara, Tong Lu, Trung Quy Phan, Chew Lim Tan, Graphics and Scene Text Classification in Video, *The 22nd International Conference on Pattern Recognition*, pp. 4714-4719, 2014.
- [7] Hrishikesh B. Aradhye, Gregory K. Myers, Exploiting Videotext "Events" for Improved Videotext Detection, *The 9th International Conference on Document Analysis and Recognition*, pp. 894-898, 2007.
- [8] Chien-Cheng Lee, Yu-Chun Chiang, Huang-Ming Huang, Chun-Li Tsai, A Fast Caption Localization and Detection for News Videos, *The 2nd International Conference on Innovative Computing Information and Control*, pp. 226-229, 2007.
- [9] Toshio Sato, Takeo Kanade, Ellen K. Hughes, Michael A. Smith, Shin'ichi Satoh, Video OCR: Indexing Digital News Libraries by Recognition of Superimposed Captions, *Springer Multimedia Systems*, Vol. 7, Iss. 5, pp. 385-395, September 1999.
- [10] Johann Poignant, Laurent Besacier, Georges Quenot, Franck Thollard, From Text Detection in Videos to Person Identification, *IEEE International Conference on Multimedia and Expo*, pp. 854-859.
- [11] Zhe Yang, Ping Shi, Caption Detection and Text Recognition in News Video, *The 5th International Congress on Image and Signal Processing*, pp. 188-191, 2012.
- [12] Huiping Li, David Doermann, Omid Kia, Automatic Text Detection and Tracking in Digital Video, *IEEE Transactions on Image Processing*, Vol. 9, No. 1, pp. 147-156, January 2000.
- [13] Rainer Lienhart, Axel Wernicke, Localizing and Segmenting Text in Image and Videos, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 12, No. 4, pp. 256-268, April 2012.
- [14] Michael R. Lyu, Jiqiang Song, Min Cai, A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction, *IEEE Transactions on Circuits and Systems For Video Technology*, Vol. 15, No. 2, pp. 243-255, February 2005.
- [15] Xiaou Tang, Xinbo Gao, Jianzhuang Liu, Hongjiang Zhang, A Spatial-Temporal Approach for Video Caption Detection and Recognition, *IEEE Transactions on Neural Networks*, Vol. 13, No. 4, pp. 961-971, July 2012.
- [16] P. Shivakumara, N. V. Kumar, D. S. Guru, C. L. Tan, Separation of Graphics (Superimposed) and Scene Text in Video Frames, *The 11th International Workshop on Document Analysis Systems*, pp. 344-348,

2014.

[17] John Canny, A Computational Approach to Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, No. 6, pp. 679-698, November 1986.

[18] Xu Zhao, kai-Hsiang Lin, Yuxiao Hu, Yuncai Liu, Thomas S. Huang, Text from Corners: A Novel Approach to Detect Text and Caption in Videos, IEEE Transactions on Image Processing, Vol. 20, No. 3, March 2011.

[19] Sanghee Lee, Hansung Park, Jungil Ahn, Youngsang On, Kanghyun Jo, Overlay Text Graphic Region Extraction for Video Quality Enhancement Application, Journal of Broadcasting Engineering, Vol. 18, No. 4, pp. 559-571, July 2013.

저 자 소 개

이 상 희



- 1994년 2월 : 경북대학교 전자공학과 졸업(공학사)
- 2009년 2월 : 울산대학교 자동차선박기술대학원 메카트로닉스/IT전공 졸업(공학석사)
- 2010년 3월 ~ 현재 : 울산대학교 전기공학부 전기전자공학전공 박사과정
- 1997년 4월 ~ 현재 : (주)울산방송(ubc)
- ORCID : <http://orcid.org/0000-0002-3229-2533>
- 주관심분야 : 디지털멀티미디어, 디지털 방송시스템, 2D/3D 영상처리, 컴퓨터 비전(CV)

안 정 일



- 1988년 2월 : 경북대학교 전자공학과 졸업(공학사)
- 1991년 8월 : 경북대학교 전자공학과 졸업(공학석사)
- 2007년 3월 ~ 현재 : 경북대학교 전자공학과 박사과정
- 1991년 11월 ~ 1995년 3월 : LG전자 영상미디어 연구소 HDTV팀
- 1995년 4월 ~ 1997년 8월 : 현대방송
- 1997년 8월 ~ 현재 : (주)울산방송(ubc)
- ORCID : <http://orcid.org/0000-0001-5492-3052>
- 주관심분야 : 디지털방송시스템, 2D/3D 영상처리

조 강 현



- 1989년 : 부산대학교 정밀기계공학과 졸업(공학사)
- 1993년 : 일본 Osaka대학 대학원 전자제어기계공학전공 졸업(공학석사)
- 1997년 : 일본 Osaka대학 대학원 전자제어기계공학전공 졸업(공학박사)
- 1997년 ~ 1998년 : ETRI 시스템공학연구소 포스트 닥터 연구원
- 1998년 ~ 현재 : 울산대학교 전기공학부 전기전자공학전공 교수
- 현재 : IEEE, ICROS, IEIE, KIISE, KIPS, KSAE, KRS, KSBE 등 정회원 및 이사 역임
- ORCID : <http://orcid.org/0000-0001-8317-6092>
- 주관심분야 : 컴퓨터 비전(CV), 사람과 컴퓨터의 상호작용(HCI), 지능시스템 연구(IS)