

동영상에 삽입된 자막 내 문자영역화소추출

안권재[○], 김계영^{*}

[○]송실대학교 컴퓨터학과

^{*}송실대학교 컴퓨터학부

e-mail: ankwonjae@naver.com gykim11@ssu.ac.kr

Extraction text-region's pixel on caption of video

Kwon-Jae An[○], Gye-Young Kim^{*}

[○]Dept. of Computing, Soongsil University

^{*}School of Computing, Soongsil University

● 요약 ●

본 논문은 동영상 내 삽입된 자막을 문자인식이 가능하도록 문자영역을 이루는 화소를 추출하는 방법을 제안한다. 최초 자막영상을 통계학적 방법을 이용하여 색상극성을 결정한다. 이 후 색상극성에 따른 잡음제거 방법을 명암값기반과 형태학적기반으로 달리한다. 제안된 방법은 각 색상결정에 따른 적합한 잡음제거를 수행함으로써 추출된 화소들이 이루는 문자영역의 영상을 이용하여 문자인식을 수행하였을때 기존방법보다 높은 문자인식률을 보였다.

키워드: 문자추출(text extraction), 색상극성(color polarity), 문자인식(character recognition)

I. 서론

현재 저장매체의 발달과 전송매체의 발달로 대용량의 데이터 저장과 전송이 매우 용이하게 되었다. 이러한 발달은 멀티미디어에도 영향을 크게 미쳤으며, 특히 동영상 제작이 빈번히 일어나고 있다. 이렇게 대량으로 제작되는 동영상 색인에 대한 필요성은 꾸준히 제기되어 왔으며, 이러한 요구에 맞춰 동영상을 색인하기 위한 동영상 내 정보들을 추출하는 다양한 연구들이 수행되어왔다. 특히 동영상에서 자막영역을 찾는 연구[1, 2, 3]가 다수 수행되었다. 또한 찾아진 자막에 대하여 문자인식이 가능하도록 문자를 추출하는 연구[2, 3, 4]들도 수행되었다.

문자를 추출하는 방법 중 [3]의 문자추출방법은 기존 방법[4, 5]보다 성능이 우수하다고 알려져있다[2, 3]. 하지만 [3]의 문자추출 방법은 문자추출 후 잡음이 많고, 문자정보가 문자의 획두께를 임의적인 구간값으로 정하기 때문에 획두께가 임의적으로 정한 값에 맞지 않으면, Inward-filling 때 문자정보가 손실되는 경우가 발생하기도 한다.

본 논문에서는 이러한 문제점을 보완하기 위하여 자막에 대한 색상극성 결정 이 후, 그 결과에 따라 잡음제거 방법을 달리한다.

II. 본론

자막에 대해 문자영역을 이루는 화소만을 추출하기 위해 최초 입력자막에 대해 색상극성 결정을 수행한다. 색상극성 결정은 [7]

의 방법을 사용한다. [7]은 [6]을 이용하여 얻어진 이진화 영상에서 해당 논문에서 [그림 1]에서 보는 바와 같이 제안된 2개의 마스크를 이용하여 검은색 에지와 흰색 에지를 구한다. 이 후 구해진 각각의 색상에 대한 총합의 비율을 구한 후 문자 단위로 외곽에지를 제거 후 다시 에지들의 총합의 비율을 구하여 이전에 구해진 총합의 비율과 비교하여 색상에 대한 극성결정을 한다.

$$K_i(x, y) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad K_w(x, y) = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

그림 1. [7]에서 제안된 에지 검출 마스크

Fig 1. Edge mask proposed on [7]

이 방법은 다음 [표 1]에서 보는 것과 같이 4가지의 결과가 나온다.

표 1. [7]의 결과 및 자막 상태
 tabel 1. result of [7] and caption's status

결과	자막 상태
WonW	밝은배경에 밝은문자
WonB	어두운배경에 밝은문자
BonW	밝은배경에 어두운문자
BonB	어두운배경에 어두운문자

[3]의 문자추출 방법은 위의 색상결정 이후 색상극성의 결과 중 문자의 색상만을 고려하여 문자추출을 수행한다. 하지만 본 논문에서는 문자와 배경간의 관계를 고려하여 문자를 추출한다.

색상극성 결정 결과가 WonW, BonB와 같이 배경과 문자의 명암값 대비가 좋지 않은 경우 기존 [3]과 같이 Dam-point-labeling 방법을 수행한다.

색상극성 결정 결과가 BonW, WonB가 나온 경우는 문자와 배경간의 명암값 대비가 좋은 경우다. 이러한 경우 명암값에 대한 임계값을 새로 구하여 적용함으로써 잡음을 제거할 수 있다[8].



그림 2. 새로운 임계값을 이용한 잡음제거
 (a) 원본영상. (b) 색상극성 결정 후 영상. (c) 결과 영상
 Fig 2. Reducing noise using new threshold value
 (a) original image, (b) image after classifying color polarity, (c) result image.

하지만 이 방법은 자막 내 문자가 2가지색 이상으로 존재 할 경우 [그림 2]에서 보는 것과 같이 문자정보가 손실되는 경우가 발생할 수 있다. 그렇기 때문에 자막에서 서로 다른 색상의 문자가 존재할 경우 [그림 3]에서 보는 것과 같이 영역을 분할하여 각 영역에 대해 [8]의 잡음제거 방법을 수행한다.

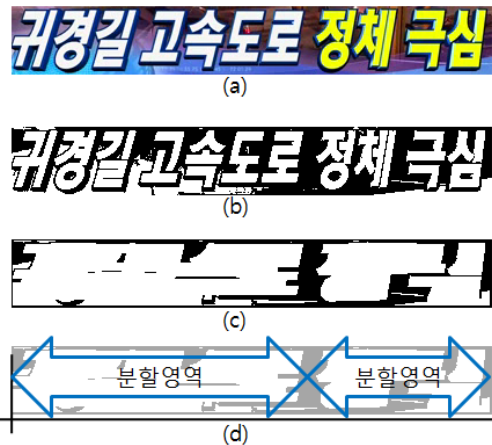


그림 3. 문자영역 분할 (a) 원본영상. (b) 색상극성 결정 후 영상.
 (c) (b)를 스미어링 한 영상. (d) (c)에 대한 문자영역 분할
 Fig 3. Splitting text region (a) original image, (b) image after classifying color polarity, (c) smearing image in (b), (d)Splitting text region in (c)

문자영역 분할 방법은 다음과 같다.

색상극성 결정 후 문자화소를 흰색으로 전환한다. 이 후 연결된 흰색화소에 대해 연결성 분석을 통해 각 객체 및 객체의 평균 명암값을 구하고 가로방향 스미어링을 이용하여 이웃 객체가 명암값이 유사할 경우 연결한다. 이후 흰색화소에 대해 세로방향 투영을 실시하여 투영된 결과에서 각 bin과 이웃 bin 대해 연결성 분석을 실시하되 누적 화소의 개수가 일정 개수보다 적은 bin이 존재할 경우 잡음 객체가 누적된 것으로 판단하고 누적화소가 없는 bin으로 판단한다. 이 후 분할영역은 잡음제거 시 누락되는 구간이 없도록 비어있는 bin의 중간 지점까지 분할영역을 잡는다.

III. 실험결과

본 논문의 실험환경의 시스템은 듀얼코어 2.8Ghz, 4G RAM이며, 구현은 MFC2005로 구현하였다. 실험은 자막에 대해 본 논문에서 제안된 방법의 문자추출방법과 Lyu의 방법의 문자추출방법으로 얻어진 영상을 문자인식을 수행하였다. 색상극성 분류 후 BonB, WonW는 기존 Lyu의 문자추출방법[3]과 같고 BonW, WonB는 다르기 때문에 BonW, WonB에 대한 자막만을 실험에 사용하였다. 실험에 쓰인 자막은 각 방송사 MBC, KBS, SBS 별로 방송하는 뉴스에서 30개씩 총 90개를 수작업으로 추출하였다. 동영상의 해상도는 640 X 360이었으며, 실험에 사용된 자막은 자막의 크기에 따라 자막의 경계에서 1개에서 10개의 화소의 여분을 두고 추출하였다. 그 이유는 [3]의 문자추출 방법이 약간의 여분공간을 필요로 하기 때문이며, 실험에서 [7]의 방법이 약간의 여분이 있을 때 정확도가 높았다. 하지만 본 논문은 문자추출방법에 관한 논문이므로 여분 공간 확보와 [7]의 정확도 관계에 대한 자세한 실험은 다루지 않았다.

다음 [표2]는 제안된 문자추출방법과 [3]의 문자추출방법에 대한 예이다.

표 2. 제안된 문자추출 방법과 Lyu의 문자추출 방법에 대한 문자추출 결과 예

table 2. Example about result of proposed text extraction method and Lyu's text extraction method

원본영상	제안된문자추출방법	Lyu[3]문자추출 방법

다음 [표 3]은 전체 자막 중 색상극성이 정확히 된 87개의 자막에 존재하는 1028개의 문자를 대상으로 한 문자인식의 결과이며, 참고로 총 90개의 자막에 존재하는 문자의 개수는 1078개였다.

표 3. 제안된 문자추출 방법과 Lyu의 문자추출 방법에 대한 문자인식 결과

table 3. Result of character recognition using proposed text extraction method and Lyu's text extraction method

	인식된 문자개수	문자인식률
제안된문자추출방법	604	58.8%
Lyu[3]문자추출 방법	558	54.3%

위의 [표 2]에서 보는 것과 같이 제안된 방법은 기존 [3]의 문자추출의 단점인 잡음이 많은 점과 문자정보가 손실되는 문제점을 어느정도 보완한 것을 볼수 있다. 또한 [표 3]에서 보는 바와 같이 제안된 문자추출방법으로 얻은 영상에 대한 문자인식 결과가 [3]의 문자추출방법으로 얻은 영상에 대한 문자인식결과 보다 우수한 것을 알 수 있었다.

IV. 결론

본 논문은 동영상 내 인위적으로 삽입된 자막에 대해 문자인식에 적합하도록 문자를 추출하는 방법에 대해 기술하였다. 실험에서 알 수 있듯이 각 색상극성에 따라 효율적인 잡음제거 방법을 채택함으로써 문자인식률을 높일 수 있었다.

향후, BonB, WonW에 대한 [3]의 방법보다 우수한 형태학적 잡음제거 방법에 대해 연구할 계획이다.

참고문헌

- [1] Zhong, Y., Zhang, H., Jain, Anil K., "Automatic caption localization in compressed video", IEEE Trans. Pattern Anal. Machine Intell. 22 (4), pp.385~392, 2000
- [2] Kim, Wonjun. Kim, Changick., "A new approach for overlay text detection and extraction from complex video scene.", IEEE transactions on image processing, a publication of the IEEE Signal Processing Society, v.18, no.2, pp.401~411, 2009
- [3] 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, a publication of the Circuits and Systems Society, v.15, no.2, pp.243~255, 2005
- [4] Rainer Lienhart, Axel Wemnicke, "Localizing and segmenting text in images, videos and web pages," IEEE transactions on circuits and systems for video technology, vol. 12, no. 4, pp.256~268, 2002
- [5] T. Sato, T. Kanade, E. K. Hughes, and M. A. Smith, "Video OCR for digital news archive" in Proc. IEEE Workshop Content-Based Access Image Video Database, 1998, pp. 52~60.
- [6] N. Otsu., "A threshold selection method from gray-level histograms", IEEE Transactions on Systems, Man and Cybernetics, vol.SMC~9, pp. 62~66, 1979
- [7] Jiqiang Song, Min Cai, Michael R.Lyu, "A ROBUST STATISTIC METHOD FOR CLASSIFYING COLOR POLARITY OF VIDEO TEXT", 2003. ICME '03. Proceedings. 2003 International Conference on Volume 2, pp.581~584, 2003
- [8] Kwon-Jae An, Gye-Young Kim, "Determining intensity value of characters and backgrounds on caption", Proceedings of the Korean Society of Computer Information Conference, vol.18 no.2, pp.125~128, 2010