해리스 코너 검출기를 이용한 비디오 자막 영역 추출

(Text Region Extraction from Videos using the Harris Corner Detector)

김 원 준 † 김 창 익 ††
(Wonjun Kim) (Changick Kim)

요 약 최근 많은 TV 영상에서 시청자의 시각적 편의와 이해를 고려하여 자막을 삽입하는 경우가 늘어나고 있다. 본 논문에서는 자막을 비디오 내 하단부에 위치하는 인위적으로 추가된 글자 영역으로 정의한다. 이러한 자막 영역의 추출은 비디오 정보 검색(video information retrieval)이나 비디오 색인(video indexing)과 같은 응용에서 글자 추출을 위한 첫 단계로 널리 쓰인다. 기존의 자막 영역 추출은 자막의 색, 자막과 배경의 밝기 대비, 에지(edge), 글자 필터 등을 이용한 방법을 사용하였다. 그러나 비디오 영상 내 자막이 갖는 낮은 해상도와 복잡한 배경으로 인해 자막 추출에 어려움이 있다. 이에 본 논문은 코너 검출기(corner detector)를 이용한 효율적인 비디오 자막 영역 추출 방법을 제안하고자 한다. 제안하는 알고리즘은 해리스 코너 검출기를 이용한 코너 맵 생성, 코너 밀도를 이용한 자막 영역 후보군 추출, 레이블링(labeling)을 이용한 최종 자막 영역 결정, 노이즈(noise) 제거 및 영역 채우기의 네 단계로 구성된다. 제안하는 알고리즘은 색 정보를 이용하지 않기 때문에 여러 가지 색으로 표현되는 자막 영역 추출에 적용 가능하며 글자 모양이 아닌 글자의 코너를 이용하기 때문에 언어의 종류에 관계없이 사용 될 수 있다. 또한 프레임간 자막 영역 업데이트를 통해 자막 영역 추출의 효율을 높였다. 다양한 영상에 대한 실험을 통해 제안하는 알고리즘이 효율적인 비디오 자막 영역 추출 방법임을 보이고자 한다.

키워드 : 자막 영역, 비디오 정보 검색, 비디오 색인, 해리스 코너 검출기

Abstract In recent years, the use of text inserted into TV contents has grown to provide viewers with better visual understanding. In this paper, video text is defined as superimposed text region located of the bottom of video. Video text extraction is the first step for video information retrieval and video indexing. Most of video text detection and extraction methods in the previous work are based on text color, contrast between text and background, edge, character filter, and so on. However, the video text extraction has big problems due to low resolution of video and complex background. To solve these problems, we propose a method to extract text from videos using the Harris corner detector. The proposed algorithm consists of four steps: corer map generation using the Harris corner detector, extraction of text candidates considering density of corners, text region determination using labeling, and post-processing. The proposed algorithm is language independent and can be applied to texts with various colors. Text region update between frames is also exploited to reduce the processing time. Experiments are performed on diverse videos to confirm the efficiency of the proposed method.

Key words : Text region, video information retrieval, video indexing, Harris corner detector

1. 서 론

디지털 멀티미디어 기술의 발전과 영상 단말기 사용 증가로 비디오는 가장 중요한 정보 매체 중 하나가 되었다. 비디오는 음성과 영상 정보를 동시에 포함하고 있는 가장 일반적인 멀티미디어 데이타라고 할 수 있다. 이러한 비디오는 대부분 사용자의 시청 편의를 위해 현

† 학생회원 : 한국정보통신대학교 전자공학과
jazznova@icu.ac.kr
†† 정 회 원 : 한국정보통신대학교 전자공학과
ckim@icu.ac.kr
논문접수 : 2007년 1월 22일
심사완료 : 2007년 5월 26일

재 콘텐츠(contents)에 대한 설명과 같은 중요한 정보를 자막을 통해 보여준다. 예를 들면, 뉴스 영상에서 중요 기사를 자막으로 보여준다거나 영화의 경우 등장 인물의 대사를 자막으로 보여준다. 또한, 스포츠 경기에서 선수의 이름과 현재 점수 상황 등을 자막을 통해 보여준다[1]. 이와 같은 자막은 일반적으로 장면 자막(scene text)과 인위적 자막(artificial text)으로 나눌 수 있다[2]. 장면 자막은 광고판 등과 같이 배경으로써 영상의 내용과 거의 관계없이 나타나지만 인위적 자막은 뉴스의 헤드라인(headline)과 같이 해당 기사와 관련하여 정확한 정보를 보여준다. 따라서 인위적 자막은 비디오 색인이나 정보 검색에 유용하게 사용될 수 있다[3]. 추출된 자막의 확대를 통해 소형 이동형 단말기에서도 쉽게 영상에 대한 정보를 얻을 수 있으며 시각적 장애가 있는 사람들을 위한 정보를 제공 할 수도 있다[4]. 또한 자동차 번호판 분석이나 교통 표지판 분석과 같은 분야[5]에도 응용 될 수 있으므로 비디오 내 자막 영역을 효율적으로 추출할 필요가 있다. 그러나 비디오 내 자막이 갖는 낮은 해상도와 언어에 따라 다른 자막의 특성, 다양한 색으로 이루어진 복잡한 배경이 자막 추출에 어려움을 주고 있다.

이러한 문제를 극복하기 위해 기존의 연구는 주로 자막의 색 정보와 에지 정보를 이용하여 이루어졌다[1, 6-18]. L. Agnihotri[1] 등은 비디오 내 자막은 동일한 색깔을 가지고 있다고 가정하고, Red 채널을 이용하여 높은 밝기 대비를 가지는 자막의 edge를 얻는다. X. S. Hua[6] 등은 높은 밝기 대비를 가지는 프레임과 블록을 이용하여 자막을 추출한다. 그러나 영상 압축으로 인한 영상의 열화로 비디오 내 자막이 동일한 색을 가지기 어려우며, 복잡한 배경으로 인해 글자와 배경간의 밝기 대비가 높지 않은 문제점이 있다. 이와 달리[7-13]에서는 상대적으로 색보다 환경에 덜 민감한 에지를 이용하여 자막 추출을 시도하였다. M. R. Lyu[7] 등은 에지의 강도를 이용한 에지 맵을 자막 추출에 이용한다. 또한 국부적 임계값(local thresholding)을 이용하여 복잡한 배경을 제거하고 다양한 언어와 자막 크기에 적용 가능하도록 다단계 해상도를 이용한 추출 방법을 사용한다. C. Liu[11] 등은 자막의 특성을 고려한 수직, 수평, 대각선 방향의 에지 맵과 이를 통한 K 평균 군집화(K-means clustering)를 이용하여 자막을 추출한다. Y. Liu[12] 등은 자기 적응 임계값(self-adaptive thresholding)을 통해 향상된 케니 에지 검출기(improved Canny edge detector)와 직선 벡터 그래프(line vector graph)를 이용하여 자막을 추출한다. 또한 C. Mi[13] 등은 [12]에서 제안하는 방법을 바탕으로 프레임 누적을 통해 자막 영역을 추출한다. 그러나 여전히 복잡한 배경에서 많은 에지가 자막 영역과 함께 검출되는 문제점이 있다. 그 밖에도 T. Sato[14] 등이 제안한 영문자의 특성을 반영한 수직, 수평, 대각선 방향의 필터를 이용하여 자막을 추출하는 방법, B. T. Chun[15] 등이 제안한 자막의 위상 특징과 색 정보를 동시에 이용하는 방법, H. Li[16] 등이 제안한 자동 신경 회로망 학습(automated neural network training)을 기반으로 한 자막 추출방법 등이 있으나, 이러한 방법들 역시 군집화 과정이나 사용자에 의한 선행 동작 필요, 여러 프레임을 사용해야 하는 문제점을 가지고 있다.

기존의 연구들과 달리 본 논문에서는 자막이 위치하는 영역에는 에지들의 교차점들이 밀집하여 나타난다는 사실에 착안하여 코너 검출기(corner detector)를 이용한 자막의 색과 언어의 특성에 관계없이 효율적으로 자막 영역을 추출하는 방법을 제시하고자 한다. 논문의 구성은 다음과 같다. 먼저, 2장에서 해리스 코너 검출기를 이용한 자막 추출 방법을 제안한다. 3장에서는 다양한 비디오에 대한 실험결과를 제시하며 4장에서 결론 및 향후 과제를 논의한다.

2. 제안하는 알고리즘

먼저 픽셀의 집합 R을 다음과 같이 정의한다. $R=\{(x,y)\,;1\le x\le X,\ 1\le y\le Y\}$이고, 여기서 영상의 크기는 $X\times Y$이다. 우리의 목표는 입력 프레임의 매 프레임을 자막 영역과(Text region) 그 외의 영역(background)으로 분할하는 것이다. P를 R의 분할된 한 형태로 정의하면 다음과 같이 나타낼 수 있다. $P=\{TR,BG\}$. 여기서 TR은 자막 영역, BG는 그 외의 배경 영역을 나타낸다. 입력 비디오에서 n번째 프레임의 자막 영역은 TR_n으로 나타내기로 한다.

2.1 코너 맵 생성

에지(edge)는 영상 내 자막 영역 이외에도 많이 존재할 수 있으나 코너는 각이 진 부분에서 주로 발생하기 때문에 상대적으로 자막 영역에 집중적으로 분포한다. 따라서 영상 내에 존재하는 코너의 밀도를 이용하여 효율적으로 자막을 추출 할 수 있다. 잘못된 코너 추출을 줄이기 위해 코너 맵을 생성하기 전 가우시안 필터를 사용한다. 본 논문에서는 특징 점 추출에 널리 쓰이는 해리스 코너 검출기를 이용하여 이러한 코너 맵을 생성한다. 해리스 코너 검출기에 대하여 간단히 살펴보면, 기본적으로 지역적인 신호 변화를 측정할 수 있는 local auto-correlation 함수에 바탕을 두고 있다. 영상 내에서 점 (x,y)가 주어지고 이에 대한 변화량이 $(\Delta x,\Delta y)$로 주어지면 auto-correlation 함수는 다음과 같이 표현할 수 있다[19].

$$c(x,y)=\sum_{W}[I(x_i,y_i)-I(x_i+\Delta x,y_i+\Delta y)]^2 \quad (1)$$

$I(\cdot,\cdot)$은 밝기를 나타내며 (x_i,y_i)은 가우시안 윈도우 W 내부의 점들을 나타낸다. $(\Delta x,\Delta y)$만큼 움직인 영역을 테일러 확장을 이용하여 표현하면 아래와 같다.

$$I(x_i+\Delta x,y_i+\Delta y)\approx I(x_i,y_i)+[I_x(x_i,y_i)\ I_y(x_i,y_i)]\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix} \quad (2)$$

$I_x(\cdot,\cdot)$, $I_y(\cdot,\cdot)$은 각각 x, y에 대한 그레디언트(gradient)를 나타낸다. 식 (2)를 식 (1)에 대입하여 정리하면,

$$c(x,y)=[\Delta x\ \Delta y]\begin{bmatrix}\sum_W (I_x(x_i,y_i))^2 & \sum_W I_x(x_i,y_i)I_y(x_i,y_i)\\ \sum_W I_x(x_i,y_i)I_y(x_i,y_i) & \sum_W (I_y(x_i,y_i))^2\end{bmatrix}\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix} \quad (3)$$

이고, 식 (3)의 두 번째 행렬을 $C(x,y)$이라 하고 이를 이용하여 코너의 정도를 판단하게 된다. 코너의 정도는 식 (4)에 의하여 결정되고 각 픽셀에 대하여 코너의 정도를 나타내는 O값을 이용하여 코너 맵을 생성한다 [20].

$$O(x,y)=\det(C(x,y))-k[trace(C(x,y))]^2 \quad (4)$$

$$HCM(x,y)=\begin{cases}1 & if\ O(x,y)>0\\ 0 & otherwise\end{cases} \quad (5)$$

여기서 k=0.04를 사용하였으며 HCM은 O값을 이용하여 이진화한 코너 맵을 나타낸다. 그림 1은 뉴스 이미지에 대한 코너 맵 생성 결과를 보이고 있다. 에지 맵을 이용하였을 때 보다 코너 맵을 이용하였을 때의 장점을 보이기 위해 우리는 각각의 맵에서 자막 영역으로 검출되는 픽셀의 비율을 다음과 같이 정의한다.

$$A_{edge}=\frac{Card(TE_n)}{Card(E_n)},\quad A_{corner}=\frac{Card(TC_n)}{Card(C_n)} \quad (6)$$

여기서 $Card(E_n)$과 $Card(C_n)$은 에지와 코너로 검출된 총 픽셀 수를 나타낸다. $Card(TE_n)$과 $Card(TC_n)$은 각각 에지 맵과 코너 맵에서 자막 영역으로 검출된 픽셀 수를 나타낸다. 그림 2에서 A_{edge} 값은 0.262211이고 A_{corner} 값은 0.635269이다. 이러한 결과는 기존 연구에서 널리 쓰이던 에지 맵보다 코너 맵을 이용하는 것이 자막 영역 검출에 더 효율적임을 보여주고 있다.

2.2 코너 밀도를 이용한 자막 후보군 추출

코너를 이용하여 자막 후보군을 추출하기 위해서 블록 가중치(block weight)를 이용한다. 이를 위해 식 (5)에서 구한 이진화된 코너 맵을 사용한다. 글자의 크기는 보통 8~24 픽셀 정도이기 때문에[7] 최소 글자 크기와 노이즈로 인한 차이를 고려하여 10 픽셀×10 픽셀 크기의 블록을 사용하였다. 블록을 이용하여 영상을 스캔하는 과정에서 글자가 블록과 정확히 정합되지 않을 수 있기 때문에 수직, 수평 방향으로 블록 크기의 반인 5 픽셀씩 이동하면서 스캔하도록 하였다. 블록 안의 코너 비중이 블록 크기의 30%이상 일 때 자막 영역이라고 판단하고 해당 블록 전체를 1값으로 채워준다. 블록 전체를 채워주는 것은 영상의 열화로 글자임에도 불구하고 코너가 추출 되지 않은 부분에 대하여 침식이 일어나지 않도록 하기 위함이다. 식으로 표현하면 다음과 같다.

$$corner\_density=\frac{\sum_{(x,y)\in B} HCM(x,y)}{Block\_size}$$

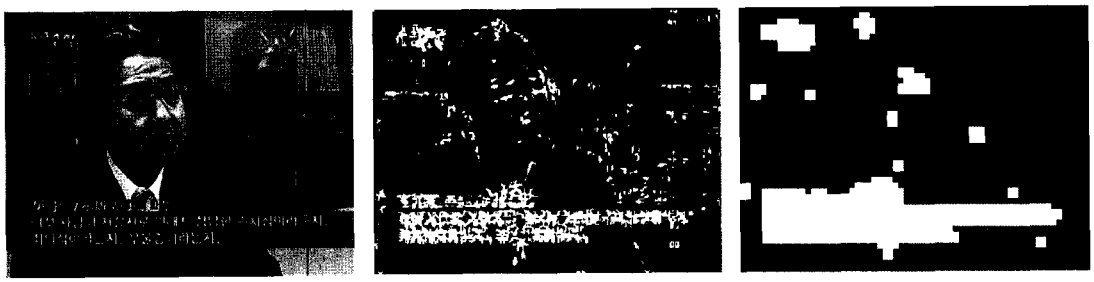
$$HCM(x,y)=\begin{cases}1, & if(corner\_density>0.3)\\ 0, & otherwise\end{cases} \quad (7)$$

여기서 B는 현재 블록을 말한다. 이를 이용하여 자막 후보군을 추출한 결과를 그림 2에 나타내었다.

그림 2(c)를 보면 블록 가중치를 이용하여 자막 영역

(a) (b) (c)

그림 1 (a) 뉴스 영상. (b) Sobel 마스크를 이용한 에지 맵. (c) 해리스 코너 검출기를 이용한 코너 맵. 코너가 자막 영역 안에 밀집되어 분포함을 알 수 있다.



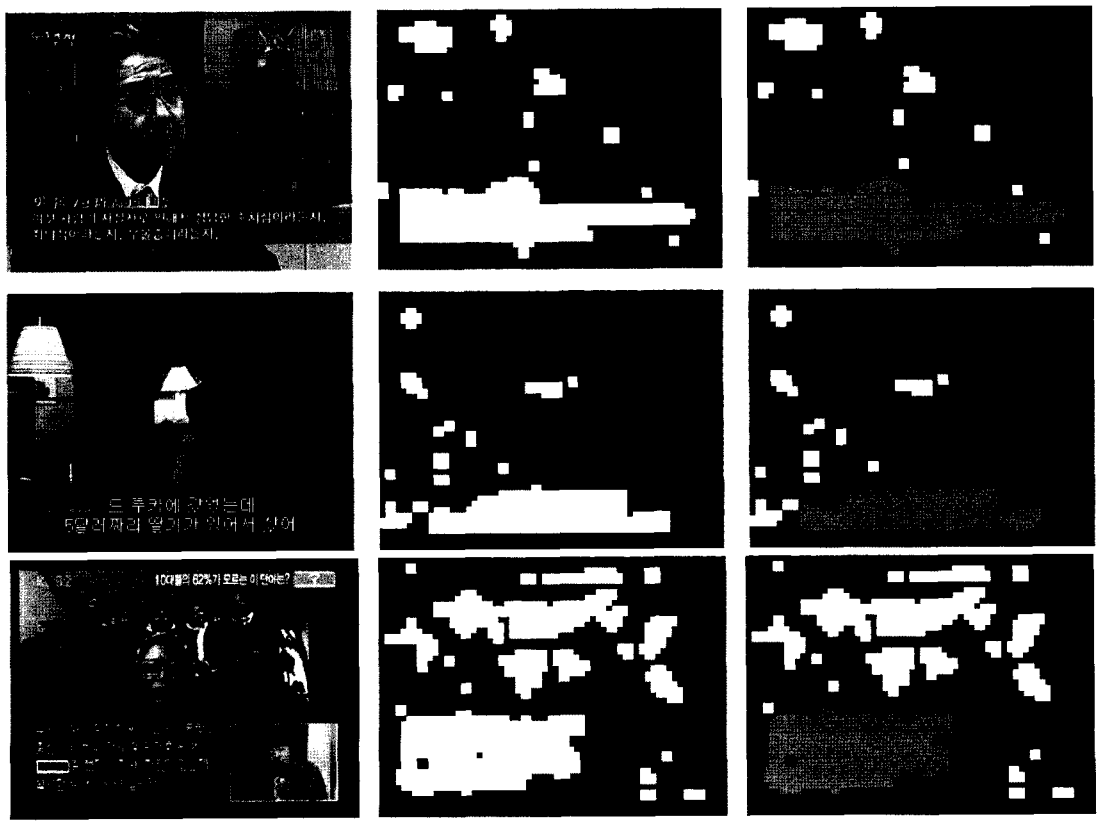
(a) (b) (c)

그림 2 (a) 뉴스 영상. (b) 이진화된 코너 맵. (c) 블록 가중치를 이용하여 추출한 자막 후보군

이 아닌 곳에서 나타난 코너가 많이 사라졌음을 알 수 있다.

2.3 레이블링을 이용한 최종 자막 영역 결정

본 논문에서는 빠른 속도를 위해서 스택(stack)을 이용한 레이블링 기법을 사용하였다. 이는 픽셀의 연결성(connectivity)에 따라 영역을 구분하기 위해 추출된 자막 후보군 중에서 1값을 갖는 픽셀의 이웃 픽셀 값을 조사한다. 같은 값의 픽셀이 나타나면 현재 픽셀에 레이블을 할당하고 현재의 위치와 이동 할 방향을 스택을 이용하여 저장한 후 1값을 갖는 새로운 픽셀로 이동하면서 영역을 구분해 나가는 방법이다. 레이블링 과정에서 과도한 자기 호출로 시스템 스택이 넘치는 것을 방지하고 고속 동작을 위해 사용자가 스택의 크기를 정의할 수 있도록 하였다. 다양한 영상에서 이진화된 코너 맵에 블록 가중치를 적용한 후 자막 영역에 대하여 레이블링을 수행하면 그림 3과 같다. 레이블링이 끝난 후,



(a) (b) (c)

그림 3 (a) 자막이 삽입 된 다양한 영상들. (b) 블록 가중치를 이용하여 추출한 자막 후보군. (c) 최종 자막 영역 추출

영상 내에 존재하는 모든 영역 중 최소 자막의 크기를 고려하여 일정 크기 이상인 영역에 대해서 크기가 가장 큰 두 개의 영역을 선정하여 그 중 하단에 위치한 영역을 최종 자막 영역으로 결정한다. 본 논문에서는 블록의 크기와 일반적으로 자막 영역 길이가 영상 내에서 차지하는 비율을 고려하여 다음과 같은 임계값을 사용한다. 여기서 width는 입력 영상의 가로 길이를 의미한다.

$$Threshold_{label} = 10 \times (width/3) \quad (8)$$

노이즈가 자막 영역과 같은 레이블로 간주되는 것을 방지하기 위해 4 인접성(4 adjacent)을 이용하여 레이블링 한다. 복잡한 배경으로 인해 자막 영역에 근접해서 발생하는 노이즈가 자막 영역과 같은 레이블로 간주되는 경우가 있기 때문에 이러한 빈도를 되도록 줄이기 위해서 대각선 방향의 연결성을 포함하는 8 인접성(8 adjacent)을 이용한 방법보다 4 인접성을 이용한 방법을 사용한다. 또한 4 인접성을 이용하는 방법은 레이블링을 위해 살피는 주변 픽셀 수가 8 인접성을 이용하는 방법의 반이므로 속도 측면에서도 성능이 뛰어나다고 할 수 있다.

2.4 노이즈 제거 및 영역 채우기

레이블링을 통해 얻은 최종 자막 영역은 블록을 기반으로 얻어진 결과이므로 불필요한 영역을 제거할 필요가 있다. 그림 3(c)를 살펴보면, 최종 자막 영역의 윗부분과 아랫 부분에서 불필요한 영역들을 볼 수 있다. 이러한 노이즈는 수평, 수직 방향에 대하여 글자 길이에 대한 제약조건을 이용해 제거 할 수 있다. 즉, 검출된 영역에 해당하는 각 행에 대해 가로 방향으로 화면 가로 길이의 15% 이하인 행은 제거 한다. 이렇게 노이즈를 제거한 영역은 영상 열화 등의 이유로 생성되지 못한 코너로 인해 홀(hole)을 가지고 있다. 다음과 같은 영역 채우기 기법을 통해 최종적으로 자막 영역을 디스플레이 할 수 있다. 먼저 수평 방향에 대하여 홀을 채우기 위해서 왼쪽에서 오른쪽으로 처음 픽셀 값이 0이 아닌 값을 감지하면 시작점으로 지정하고, 오른쪽에서 왼쪽으로 처음 픽셀 값이 0이 아닌 값을 감지하면 끝점으로 지정하여 시작점부터 끝점까지 홀을 채운다. 수평 방향으로 홀을 채운 후, 위쪽에서 아래쪽으로, 아래쪽에서 위쪽으로 같은 방법을 이용하여 수직 방향에 대하여 홀을 채운다. 그림 4는 그림 3의 첫 번째, 세 번째 영상에 대하여 노이즈 제거 및 영역 채우기를 적용한 결과를 나타내고 있다. 그림 4(c)와 같이 자막 영역이 잘 추출되고 있음을 볼 수 있다.

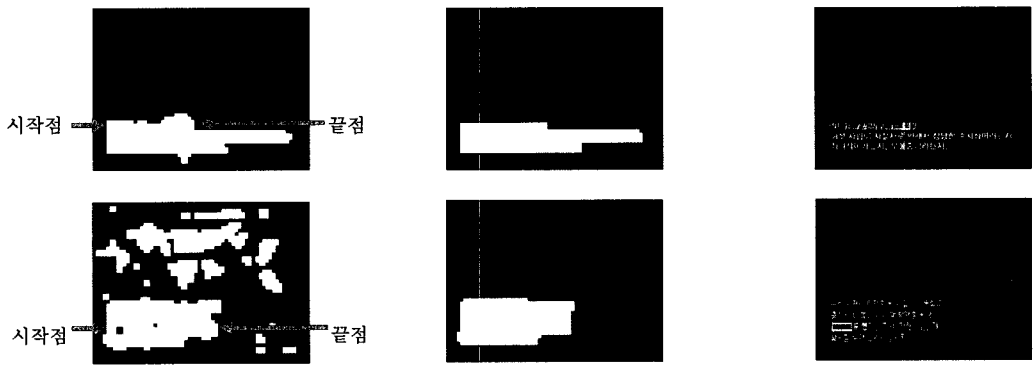
2.5 프레임간 자막 영역 업데이트

비디오 영상은 프레임간 비슷한 자막을 포함하고 있기 때문에 매 프레임마다 같은 과정을 반복 할 필요 없이 프레임간 자막 영역 업데이트를 통해 처리 과정을 줄일 수 있다. 이를 위해 프레임 전체를 이용하는 대신 빠른 처리 속도를 위해 그림 5와 같이 코너가 밀집되어 있는 R_L 영역을 이용한다. R_L 영역에서 현재 프레임과 이전 프레임의 코너 맵 차이를 구하여 임계값 보다 작으면 현재 프레임의 자막 영역에 변화가 없다고 간주하여 코너 맵 생성 이후의 과정을 무시하고 이전 자막 영역을 보여준다. 이를 식 (9)와 같이 나타낼 수 있다.

$$d(HCM_n, HCM_{n-1}) = \sum_{(x,y)\in R_L} (HCM_n(x,y) \otimes HCM_{n-1}(x,y))$$

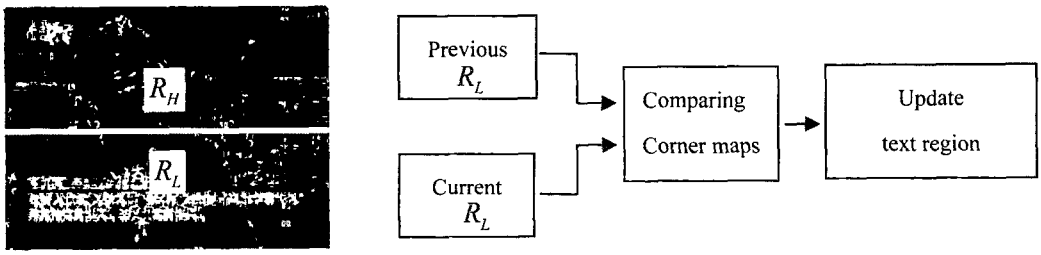
$$\begin{aligned} Compare \ : \ & if\,(d(HCM_n, HCM_{n-1}) < threshold) \\ & TR_n = TR_{n-1} \\ & otherwise, \quad find\ new\ TR_n \end{aligned} \quad (9)$$

여기서 TR_{n-1}, TR_n은 각각 이전 프레임과 현재 프레임에서 추출된 자막 영역, $d(\cdot,\cdot)$은 현재 프레임의 코너 맵과 이전 프레임의 코너 맵 차이를 나타낸다. ⓧ 연산



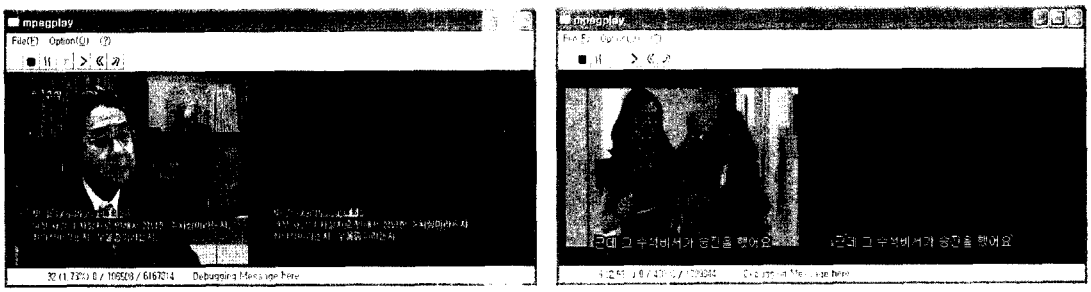
(a) (b) (c)

그림 4 (a) 레이블링을 통해 결정된 최종 자막 영역. (b) 노이즈 제거 및 영역 채우기를 수행한 후의 최종 자막 영역. (c) 추출된 실제 자막 영역



(a) (b)

그림 5 (a) R_H와 R_L. (b) 프레임간 자막 영역 업데이트 과정



(a) (b)

그림 6 실험 결과. (a) 뉴스 영상 대한 실험. (b) 영화 영상 대한 실험

그림 7 제안된 알고리즘을 다양한 영상에 적용한 결과

그림 8 글자로 이루어진 배경을 포함하여 추출되는 경우

자는 배타합(XOR)을 의미한다. 즉, 이전 프레임과 현재 프레임 코너 맵의 R_L 영역에서 같은 부분에는 0, 다른 부분에는 1을 할당한다. 이러한 프레임간 자막 영역 업데이트를 이용하여 자막 영역 추출의 효율성을 높였다.

3. 실험 결과

성능 측정을 위한 프레임 워크(framework)를 Win32 환경에서 Visual Studio 2003 (C++)을 이용하여 개발하였으며, MPEG 복호화를 위해 FFMpeg 라이브러리를 이용하였다. 실험에는 정적인(static) 영상과 동적인(dynamic) 영상을 비교하기 위해 320×240 크기의 300 프레임으로 구성된 뉴스 프로그램과 영화 영상을 사용하였다. 모든 과정은 Pentium 4 3.00GHz PC에서 실시간으로 수행되었다. 프레임간 자막 영역을 업데이트 하는 과정에서는 임계값으로 다양한 영상에 대해 실험한 결과 2000을 사용하였다.

제안된 알고리즘을 다양한 영상에 적용한 결과를 그림 7에서 보여주고 있다. 영상 내 언어, 자막의 색깔과 크기, 자막과 배경간의 밝기 대비에 관계없이 자막 영역이 잘 추출되고 있음을 알 수 있다.

그러나 글자로 이루어진 배경을 포함하고 있는 영상에서는 배경이 자막 영역으로 간주되기 때문에 정확한 자막 영역 추출이 어렵다. 그림 8은 자막 영역의 배경으로 나타나는 사람의 옷 무늬가 글자인 경우와 뉴스 영상에서 글자로 이루어진 배경이 나타났을 경우 자막 추출의 결과를 보여주고 있다. 배경에 포함된 글자가 자막 영역으로 간주되어 추출되고 있음을 볼 수 있다.

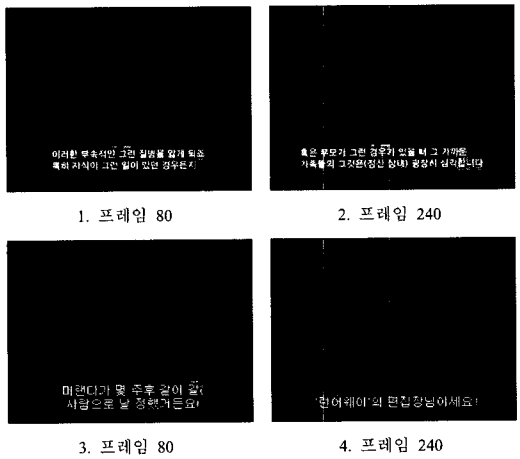
그림 6에서 보인 뉴스 영상과 영화 영상의 처리 속도에 대한 측정 결과를 표 1에 정리하였다. 프레임간 자막 영역 업데이트를 이용 할 경우 성능 향상을 알아보기 위해 프레임간 자막 영역 업데이트를 이용한 경우와 그렇지 않은 경우를 나누어서 측정 하였다.

프레임간 자막 영역 업데이트를 사용하면 코너 맵 생성 이후의 처리과정이 무시되기 때문에 처리 속도의 향상을 가져 온다. 뉴스 영상의 경우 영상 내 정적인 프레임이 대부분이기 때문에 프레임간 자막 영역 업데이트에 의해 큰 속도 향상을 가져온다. 이에 비해 영화 영상의 경우 대부분이 동적인 프레임이고 자막이 수시로 바뀌기 때문에 프레임간 자막 영역 업데이트가 뉴스 영상만큼 영향을 미치지 못한다. 따라서 정적인 프레임이 많을수록 더 큰 속도 향상을 얻을 수 있음을 알 수 있다.

추출의 정확성을 측정하기 위해 본 논문에서는 다음과 같이 정의 되는 Recall과 Precision을 사용하였다.

$$\text{Recall} = \frac{Card(TR_n \bigcap TR_{n,GT})}{Card(TR_{n,GT})}$$

$$\text{Precision} = \frac{Card(TR_n \bigcap TR_{n,GT})}{Card(TR_n)} \qquad (10)$$

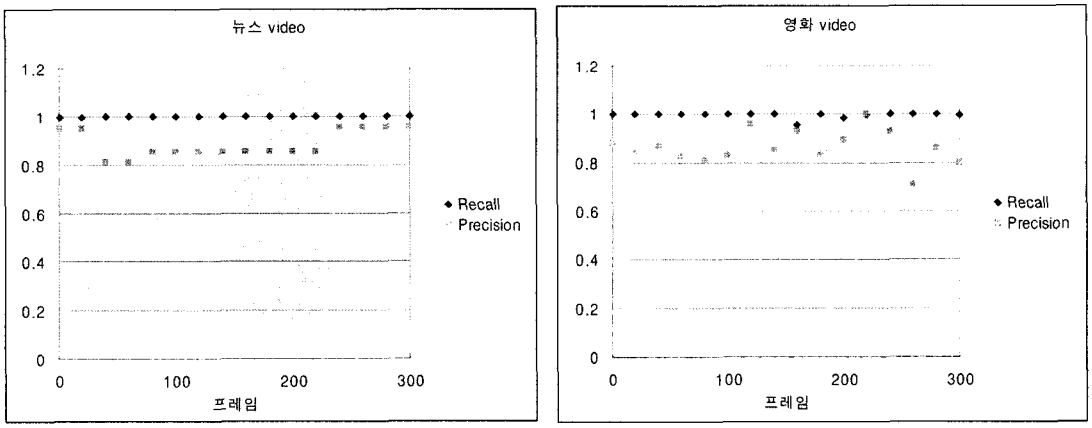


1. 프레임 80 2. 프레임 240

3. 프레임 80 4. 프레임 240

그림 9 뉴스 영상에서 자막 추출 결과(1,2)와 영화 영상에서 자막 추출 결과(3,4)

여기서 $TR_{n,GT}$ 는 n번째 영상에서의 수동으로 추출한 자막 영역(ground truth)을 나타내며, $Card(A)$ 는 영역 A 에 속하는 픽셀의 개수를 나타낸다. 각 실험 영상에 대해 20 프레임 간격으로 자막 추출한 결과를 그림 9에 나타내었고, 이에 해당하는 Recall과 Precision 값은 그림 10에 그래프로 나타내었다. 그림 10에서 보는 바와 같이 제안하는 시스템에 의해서 높은 Recall과 Precision 값을 얻을 수 있는 것을 알 수 있으며, 동적인 영화 영상에 비해 정적인 뉴스 영상의 Recall과 Precision 값이 일정하게 나타나고 있음을 알 수 있다.



(a) (b)

그림 10 20 프레임 간격으로 Recall과 Precision을 측정한 결과. (a) 뉴스 영상에 대한 결과. (b) 영화 영상에 대한 결과

4. 결론 및 토의

본 논문에서는 글자의 코너를 이용한 효율적이고 노이즈에 강건한 비디오 자막 추출 알고리즘을 제안하였다. 기존의 비디오 자막 추출 알고리즘에서 언어적 특성이나 자막의 크기, 색 등을 고려하는 것과 달리 제안하는 알고리즘은 글자의 코너를 이용하여 자막을 추출하기 때문에 언어적 특성이나 글자의 크기, 색깔 등에 대한 고려 없이 적용이 가능하다. 블록 가중치와 레이블링을 이용하여 최종 자막 영역을 추출하기 때문에 노이즈에 강건한 동작이 가능하다. 또한 프레임간 자막 영역 업데이트를 이용하여 불필요한 연산 과정을 줄여줌으로써 처리 속도를 증가시켰다.

3장의 실험 결과에서 알 수 있듯이 제안하는 알고리즘은 일반 PC에서 실시간으로 동작 가능하기 때문에 비디오 정보 검색과 같은 응용에 유용하게 사용될 수 있다. 그러나 그림 8에서와 같이 자막의 배경이 글자를 포함하고 있는 경우, 배경이 자막 영역으로 간주되어 정확한 자막 영역 추출에 어려움이 있다. 따라서 위와 같은 문제점을 해결하기 위해 추출된 자막 영역에 대하여 색상 변화 빈도수나 자막 영역의 색 정보를 추가로 이용하여 글자를 포함하고 있는 복잡한 배경에서도 정확히 자막 영역을 추출할 수 있는 강건한 알고리즘을 개발 중에 있다. 또한 시청자 편의를 위해서 자막 영역을 확대하여 볼 수 있는 기능을 추가할 계획이다. 이를 바탕으로 이동형 단말기를 위한 응용에도 사용될 수 있을 것으로 기대된다. 최종적으로 비디오 OCR (Optical Character Recognition) 응용을 위해서 추출된 자막 영역으로부터 자막을 분리하는 이진화 알고리즘을 개발 중에 있다.

참 고 문 헌

[1] L. Agnihotri and N. Dimitrova, "Text detection for video analysis," *IEEE International Workshop on Content-Based Access of Image and Video Libraries*, pp. 109-113, June 1999.

[2] J. Gllavata, R. Ewerth, and B. Freisleben, "Text detection in images based on unsupervised classification of high-frequency wavelet coefficients," *International Conference on Pattern Recognition*, vol. 1, pp. 425-428, Aug. 2004.

[3] J. Cho, S. Jeong, and B. Choi, "News video retrieval using automatic indexing of Korean closed-caption," *Lecture Notes in Computer Science*, vol. 3683, pp. 694-703, Aug. 2005.

[4] N. Ezaki, K. Kiyota, B. T. Minh, M. Bulacu, and L. Schomaker, "Improved text-detection methods for a camera-based text reading system for blind persons," *International Conference on Document Analysis and Recognition*, vol. 1, pp. 257-261, Sept. 2005.

표 1 뉴스와 영화 영상에 대한 성능 평가

	뉴스 영상	영화 영상
총 길이	300 프레임	300 프레임
자막 영역 업데이트를 이용하지 않은 경우 평균 재생 속도	18.032 frame/sec	19.506 frame/sec
자막 영역 업데이트를 이용한 경우 평균 재생 속도	22.646 frame/sec	22. 669 frame/sec
평균 재생 속도 증가율	25.59 %	16.22 %

[5] W. Wu, X. Chen , and J. Yang, "Detection of text on road signs from video," *IEEE Transaction on Intelligent Transportation Systems*, vol. 6, no. 4, pp. 378-390, Dec. 2005.

[6] X. S. Hua, P. Yin, and H. J. Zhang, "Efficient video text recognition using multiple frame integration," *International Conference on Image Processing*, vol. 2, pp. 397-400, Sept. 2002.

[7] M. R. Lyu, J. Song, and M. Cai, "A comprehensive method for multilingual video text detection, localization, and extraction," *IEEE Transaction on Circuit and Systems for Video Technology*, vol. 15, no. 2, pp. 243-255, Feb. 2005.

[8] M. Cai,, J. Song, and M. R. Lyu, "A new approach for video text detection," *International Conference on Image Processing*, vol. 1, pp. 117-120, Sept. 2002.

[9] J. Gllavata, R. Ewerth, and B. Freisleben, "A robust algorithm for text detection in images," *International Symposium on Image and Signal Processing and Analysis*, vol. 2, pp. 611-616, Sept. 2003.

[10] A. Ekin, "Local information based overlaid text detection by classifier fusion," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 753-756, May 2006.

[11] C. Liu, C. Wang, and R. Dai, "Text detection in images based on unsupervised classification of edge-based features," *International Conference on Document Analysis and Recognition*, vol. 2, pp. 610-614, Sept. 2005.

[12] Y. Liu, H.Lu, X. Xue, and Y. P. Tan, "Effective video text detection using line features," *International Conference on Control, Automation, Robotics and Vision*, vol. 2, pp. 1528-1532, Dec. 2004.

[13] C. Mi, Y. Xu, and X. Xue, "A novel video text extraction approach based on multiple frames," *International Conference on Information, Communication and Signal Processing*, pp. 678-682, Dec. 2005.

[14] T. Sato, T. Kanade, E. K. Hughes, and M. A. Smith, "Video OCR for digital news archive," *IEEE International Workshop on Content-Based Access of Image and Video Libraries*, pp.52-60, Jan. 1998.

[15] B. T. Chun, Y. Bae, and T. Y. Kim, "Caption segmentation method in videos using isodata clustering of topographical features," *IEEE Region 10 Conference TENCON*, vol. 2, pp.915-918, Sept. 1999.

[16] H. Li and D. Doermann, "A video text detection system based on automated training," *International Conference on Pattern Recognition*, vol. 2, pp. 223-226, Sept. 2000.

[17] C. Garcia and X. Apostolidis, "Text detection and segmentation in complex color images," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 2326-2329, June 2000.

[18] V. Wu, R. Manmatha, and E. M. Riseman, "Textfinder : and automatic system to detect and recognize text in images," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 21, no. 11, pp. 1224-1229, Nov. 1999.

[19] http://www.cse.yorku.ca/~kosta/CompVis_Notes/harris_detector.pdf.

[20] F. Faille, "A fast method to improve the stability of interest point detection under illumination changes," *International Conference on Image Processing*, vol. 4, pp. 2673-2676, Oct. 2004.

김 원 준

2006년 8월 서강대학교 전자공학과(학사). 2006년 8월~현재 한국정보통신대학교(ICU) 공학부 석사과정. 관심분야는 Object segmentation, Intelligent Display, Region of interest (ROI)

김 창 익

1989년 2월 연세대학교 전기공학과(학사) 1991년 2월 포항공과대학교(POSTECH) 전자전기공학과(석사). 1991년 1월~1997년 7월 SKC Ltd. R&D 센터 선임 연구원. 2000년 12월 워싱턴주립대학교 전기공학과(박사). 2000년 12월~2005년 1월 Senior Member of Technical Staff, Epson Palo Alto Laboratory, Epson R&D Inc. 2005년 2월~현재 한국정보통신대학교(ICU) 공학부 조교수