레이블링 기법과 밝기값 변화에 기반한 컬러영상의 문자영역 추출 방법

Text Area Extraction Method for Color Images Based on Labeling and Gradient Difference Method

원종길*, 김혜영*, 조진수**

경원대학교 일반대학원 전자계산학과*, 경원대학교 IT대학 컴퓨터공학과**

Jong-Kil Won(developerjk@nate.com)*, Hye-Young Kim(hyri87rumaya@nate.com)*,
Jin-Soo Cho(jscho@kyungwon.ac.kr)**

요약

영상 입출력 장치 사용이 증가함에 따라 컬러영상 내 문자영역 추출의 중요성 또한 높아지고 있다. 본 논문은 이러한 영상 내 문자영역을 효과적으로 추출하기 위해 레이블링 기법과 화소 단위의 밝기값 변화에 기반한 문자영역 추출 방법을 제안한다. 제안하는 방법은 레이블링 및 필터링 과정을 통해 비문자 영역을 미리 제거하고, 밝기값의 변화가 큰 문자영역의 특성을 이용하여 문자영역 후보군을 추출한 후 노이즈 제거 및 문자영역 병합의 후처리 과정을 통해 문자영역을 추출한다. 제안한 방법의 강점은 기존 방법보다 단순하면서도 높은 정확성에 있다. 실험 결과 제안한 방법의 정확도와 재현율, 비문자 추출의 역 비율(IRNTE)은 각각 99.59%, 98.65%, 82.30%로 측정되었다.

■ 중심어 : | 레이블링 | 컬러영상 | 문자 추출 | 밝기값 변화 |

Abstract

As the use of image input and output devices increases, the importance of extracting text area in color images is also increasing. In this paper, in order to extract text area of the images efficiently, we present a text area extraction method for color images based on labeling and gradient difference method. The proposed method first eliminates non-text area using the processes of labeling and filtering. After generating the candidates of text area by using the property that is high gradient difference in text area, text area is extracted using the post-processing of noise removal and text area merging. The benefits of the proposed method are its simplicity and high accuracy that is better than the conventional methods. Experimental results show that precision, recall and inverse ratio of non-text extraction (IRNTE) of the proposed method are 99.59%, 98.65% and 82.30%, respectively.

■ keyword : | Labeling | Color Image | Text Extraction | Gradient Difference |

* 이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (2011-0005707)

* 이 논문은 2011년도 경원대학교 지원에 의한 결과임

접수번호 : #111018-006

접수일자 : 2011년 10월 18일

심사완료일 : 2011년 11월 14일

교신저자 : 조진수, e-mail : jscho@kyungwon.ac.kr

I. 서론

최근 스캐너, 디지털 카메라 등을 포함하는 영상 입력 장치의 대중화에 따라 인쇄된 문서에서부터 전자 문서 혹은 동영상 등의 복잡한 컬러영상까지 문자영역을 추출하고 추출된 문자를 인식하는 연구의 필요성이 증가하고 있다. 문자영역 추출은 문자와 그림, 표 등과 같은 비문자를 분리하는 것으로 특히 광학 문자 인식(OCR) 기술의 보편적인 실용화를 위하여 반드시 선행되어야 하며, 차량 번호판 인식, 내용기반 문서영상 검색 시스템, 다계층 문서영상 압축(multi-layer document image compression), 비디오 인덱싱 등의 멀티미디어 시스템이나 시각장애인을 위한 주행 보조 시스템 및 콘텐츠 기반의 교육 시스템 등 다양한 분야에서 활용될 수 있다.

기존의 문자영역 추출에 대한 방법으로 화소 단위의 밝기값 변화를 이용하는 방법이 있다[1-3]. 이 방법은 문자와 배경 간 밝기 차이가 큰 문자영역의 특성과 비교적 이웃한 화소의 밝기값 차이가 적은 비문자 영역의 특성을 이용한다. 그러나 문자추출 과정에서 필요한 임계값(threshold)을 고정된 값으로 설정하거나, 표와 그래프, 다이어그램 등과 같이 밝기값 변화가 큰 비문자 영역을 문자영역으로 잘못 추출하는 등 다양한 유형의 컬러영상에 대하여 일관성 있게 좋은 성능을 보이지 못한다는 단점을 가지고 있다. 또한 연결요소와 미디언 필터(median filter)를 기반으로 문자와 비문자를 분류하는 방법이 있다[4][5]. 이는 미디언 필터를 이용하여 문자와 그림을 분류하고 영역확장 필터와 몇 가지 제한 조건들을 이용하여 잘못 분류된 요소들을 수정하는 방법을 이용하였다. 그러나 밝기값 변화를 이용하는 방법과 마찬가지로 고정된 임계값들을 사용하여 다양한 종류의 영상의 문자영역 추출에 효과적이지 못하다는 단점이 있다. 그 외 문자영역 추출 방법들로는 신경망을 이용한 학습방법[6-9]과 윤곽선(edge)을 이용한 방법[10-12], 색상 군집화(color clustering)를 이용한 방법[13-15], 텍스처를 이용한 방식과 연결성분을 이용하는 방식을 결합한 방법[16][17] 등이 있으나, 학습을 위한 전처리 단계의 번거로움이나 문자 픽셀과 비문자 픽셀

을 분류하기 위한 임계값 설정, 경계선에서의 다양한 색상변화, 문자추출 시 많은 계산량 등의 문제점이 있다.

본 논문에서는 레이블링 기법과 밝기값 변화에 기반하여 문자와 그림, 표, 다이어그램, 그래프 등이 혼합된 컬러영상에 대한 문자영역 추출 방법을 제안한다. 기존 방법과의 차이점 및 장점은 첫 번째로 간단하면서도 강한 레이블 필터링 과정을 통해 문자영역으로 추출될 수 있는 비문자 영역을 미리 제거하여 잘못 추출되는 비문자 영역을 최소화 할 수 있다. 두 번째로는 영상 내 문자의 크기나 이진화 등에 대한 임계값을 동적으로 계산하여 특정한 영상에 제한적이지 않고 다양한 컬러영상에 적용할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 제안하는 방법의 전반적인 원리와 문자영역을 추출하는 데 필요한 레이블링, 레이블 필터링, 문자영역 후보군 설정, 후처리 및 문자 레이블 결정의 5가지 세부단계에 대하여 설명한다. 3장에서는 제안한 방법의 실험결과에 대한 분석과 기존연구와의 비교를 통해 성능 분석을 하며, 4장에서는 결론 및 향후 연구계획에 대하여 기술한다.

II. 문자영역 추출 방법

문자영역 추출 기술은 광학 문자 인식의 선행 단계로써 멀티미디어 시스템이나 보조공학장치 등의 다양한 분야에서 활용될 수 있다. 제안하는 방법은 [그림 1]에서와 같이 컬러변환 및 이진화를 통한 레이블링 과정과 레이블의 특성을 이용하여 문자 및 비문자로 분류하는 레이블 필터링, 화소 단위의 밝기값의 차이를 이용하는 문자영역 후보군 설정, 노이즈 제거 및 문자 블록 설정을 포함하는 후처리, 마지막으로 문자 레이블 결정 단계로 구성된다.

본 논문에서 제안하는 방법은 영상 내 개체를 분류할 수 있는 레이블링(labeling) 기법과 화소 단위의 밝기값 변화를 이용하여 문자영역을 추출하는 방법에 기반한다. 레이블링을 통해 영상 내 각각의 개체에 대한 폭, 높이, 면적, 밀도 등의 정보를 추출할 수 있으며, 이러한 레이블 정보는 문자와 영상을 분류할 수 있는 기준이

될 수 있다. 영상 내 그래프, 다이어그램, 표, 바운딩 박스(예를 들어 글상자) 등이 존재하거나 문자영역이 비문자 영역과 근접하게 위치할 때 밝기값 변화만을 이용할 경우 문자로 잘못 인식될 수 있다. 따라서 레이블 정보를 이용하여 비문자 개체를 미리 제거함으로써 문자영역 추출의 정확성을 높일 수 있다.

그림 1. 제안하는 문자영역 추출 방법의 세부단계

화소의 밝기값 변화를 이용하는 방법[1-3]의 원리는 [그림 2]와 같이 문자영역과 비문자 영역에서 밝기값 변화의 특성을 이용하는 것이다. [그림 2]의 (a)는 문자영역과 그림영역이 혼재되어 있는 1243x444 해상도의 원본영상이고, (b)는 (a) 영상을 그레이 스케일로 변환한 후 (a)에서 청색 줄로 표시된 부분의 밝기 값을 나타내고 있으며, (c)는 (b)에서 측정된 각 위치의 화소값과 이웃한 화소값의 차이를 계산하여 나타낸 것이다. (c)에서 나타난 바와 같이 문자영역은 배경과 문자간의 급격한 밝기값 차이가 나타나며, 비문자 영역은 기준 화소와 이웃한 화소들 간 밝기값 차이가 비교적 적다. 이러한 특성을 이용하여 문자영역을 추출하는 방법은 효과적일 수 있다.

(a) 원본 컬러영상

(b) (a)에서 290 line의 각 화소 밝기값

(c) (a)에서 290 line의 각 화소 밝기값 변화

그림 2. 문자영역과 그림영역의 밝기 특성

다음은 제안하는 문자영역 추출 방법의 각 단계에 대하여 설명한다.

1. 레이블링

문자영역 추출 방법의 첫 번째 단계로는 컬러영상을 그레이 스케일 영상 및 이진 영상으로 변환하고 레이블링 과정을 수행하는 것이다. 레이블링은 연결된 화소를 하나의 개체, 즉 레이블로 정의함으로써 가로폭, 세로높이, 면적, 면적 대비 화소의 개수를 나타내는 밀도 등의 정보를 설정할 수 있다. 이러한 정보는 문자와 영상을 분류할 수 있는 기준이 되며, 밝기값 변화를 이용한 문자영역 후보군 추출 단계에서 문자의 잘못된 인식 문제를 최소화 할 수 있다. 레이블 정보는 제안하는 방법의 전반적인 단계에서 사용된다.

2. 레이블 필터링

레이블 필터링 단계는 문자영역 후보군을 추출하기 전 미리 비문자 영역을 분류하기 위한 과정이다. 화소단위의 밝기값 변화는 문자영역에서 급격하게 변하는 특징을 보이는데 이러한 특징은 그래프나 다이어그램,

도표, 바운딩 박스, 밝기 변화가 큰 고주파 특성이 두드러진 그림일 경우 문자로 잘못 인식될 수 있다. 따라서 이러한 비문자 영역의 특성을 이용하여 미리 비문자로 필터링하면 문자추출의 성능을 높일 수 있다.

레이블 필터링을 위하여 먼저 레이블의 크기정보를 이용한다. 일반적으로 문자와 비문자가 혼합한 컬러문서영상에서 비문자 레이블의 크기는 문자 레이블보다 가로 또는 세로 길이가 비교적 크다. 이러한 특성을 고려하여 식 (1)과 같이 비문자 레이블을 필터링 할 수 있는 가로, 세로의 각 길이에 대한 임계값을 계산한다.

$$T_l = \begin{cases} (\min l + 2avrl + \max l)/4, \\ \text{if } (avrl < (\min l + \max l)/2) \\ (avrl + \max l)/2, \text{ otherwise} \end{cases} \quad (1)$$

식 (1)에서 T_l 는 레이블의 가로 또는 세로 길이의 임계값이며, $\min l$, $\max l$, $avrl$ 은 각각 최소, 최대, 평균 길이를 의미한다. 식 (1)의 상위 항과 같이 레이블의 평균 길이 $avrl$ 이 최소 및 최대 길이 $\min l$, $\max l$ 의 평균값보다 작을 경우 임계값 T_l 을 평균값에 좀 더 근접하게 설정하기 위하여 평균 길이 $avrl$ 에 2를 곱하여 가중치를 두었으며, 그 외의 경우는 하위 항과 같이 T_l 을 평균 길이 $avrl$ 과 최대 길이 $\max l$ 의 평균값으로 설정하였다. 두 경우 모두 임계값을 평균 길이보다 높게 설정하였는데, 이는 평균 길이에 가까운 레이블일수록 문자일 확률이 높고 최대 길이에 가까운 레이블일수록 비문자일 확률이 높은 이유 때문이다.

다음으로 필터링을 위하여 레이블의 흑화소 밀도를 이용한다. [그림 3]의 (a)와 같이 그래프의 흑화소 밀도는 (b)와 같이 문자의 흑화소 밀도보다 비교적 낮게 측정되고, 그래프뿐만 아니라 바운딩 박스, 표 등의 비문자 영역에서도 유사한 특성을 나타낸다. 따라서 이러한 영상의 밀도 특성을 이용하여 레이블 필터링을 수행한다. 식 (2)는 레이블의 흑화소 밀도를 이용한 필터링을 나타낸다.

그림 3. 그래프 및 문자 레이블의 화소 밀도 특성

$$label = \begin{cases} non-text, & \text{if } (bp/ap < T_d) \\ text, & \text{otherwise} \end{cases} \quad (2)$$

여기서 bp 는 레이블 영역 내 흑화소의 수이고, ap 는 레이블 영역의 총 화소 수를 의미하며, T_d 는 문자 또는 비문자로 판별할 수 있는 임계값을 나타낸다. 문자 레이블을 필터링 하지 않고 그래프 및 표 등을 필터링 할 수 있도록 다양한 영상에서 문자와 그래프, 표 등의 흑화소 밀도를 조사하여 임계값 T_d 를 0.15로 설정하였다.

3. 문자영역 후보군 설정

문자영역 후보군 설정 단계는 영상 내 화소 단위의 밝기값 특성을 이용하여 maximum gradient difference (MGD) 영상을 생성하고[1-3], 문자가 될 수 있는 영역 후보군을 설정하는 과정이다.

먼저 그레이 스케일 영상의 줄(row) 단위로 식 (3)과 같이 $[-1, 1]$ 마스크(mask)를 이용하여 밝기값의 변화량을 계산한다.

$$G(x, y) = B(x, y) - B(x-1, y) \quad (3) \\ (1 \leq x < w, 0 \leq y < h)$$

식 (3)에서 $G(x, y)$ 는 이웃한 화소와의 밝기 변화량이고, $B(x, y)$ 는 픽셀의 밝기값이다. [그림 1]의 (c)와 같이 문자영역은 배경화소와 문자화소 간의 밝기값이 크게 차이가 나기 때문에 식 (3)의 결과 또한 극단적인 값이 나타나게 되고, 비문자 영역에서는 문자영역보다 비교적 적은 차이값을 나타낸다.

다음으로 MGD 영상을 생성한다. MGD는 $n \times 1$ 범위 내에서 가장 큰 밝기값 변화와 가장 작은 밝기값 변화

의 차이를 뜻하는 것으로써 식 (4)와 같이 계산된다.

$$MGD(x, y) = \max(G(x_i, y)) - \min(G(x_i, y)) \quad (4)$$

$$\left(0 \leq y < h, \frac{n}{2} \leq x < w - \frac{n}{2}, -\frac{n}{2} \leq i < \frac{n}{2} \right)$$

여기서 $MGD(x, y)$ 는 최대 밝기 변화량을 뜻하고, $G(x, y)$ 는 이웃한 화소와의 밝기 변화량을 의미한다. 또한 h, w 는 영상 전체의 세로, 가로를 의미하며, n 은 MGD 측정에 필요한 범위의 크기, i 는 범위 n 내에서의 위치(pixel position)를 나타낸다. 여기서 n 은 문자의 폭을 나타내는 것으로 식 (5)과 같이 계산한다.

$$NL T_f = \text{count} \left(\left(\frac{1}{NL} \sum_{j=1}^{NL} L_j W \right) < L_j W \right), \quad (5)$$

$$n = \frac{1}{NL T_f} \sum_{k=1}^{NL T_f} L_k W$$

여기서 NL 은 레이블의 총 수를 의미하고, count 는 괄호 내 조건을 만족하는 수를 계산하는 함수이며, j 는 레이블의 인덱스, $L_j W$ 는 해당 레이블의 가로길이를 뜻한다. 따라서 $NL T_f$ 는 레이블의 평균 가로길이보다 큰 레이블의 빈도수를 나타낸다. 그리고 식 (5)의 하위 항에서 k 는 레이블의 평균 가로길이보다 큰 레이블의 인덱스를 의미한다. 문서영상의 종류나 특징에 따라 문자의 크기가 다를 수 있지만, 이와 같이 레이블 정보를 이용하여 동적으로 n 을 계산함으로써 고정된 값을 사용하는 기존 방법들[1-3]의 문제점을 해결할 수 있다. 따라서 식 (4)를 이용하여 [그림 1]의 (a) 영상을 [그림 4]의 (a)와 같은 MGD 영상을 생성할 수 있고, 문자영역에서 MGD 값이 뚜렷한 것을 확인할 수 있다. 또한 MGD 영상에 이진화 과정을 거치면 [그림 4]의 (b)와 같이 나타나며, 이진화에 필요한 임계값은 식 (6)로써 계산된다.

$$T_b = \frac{1}{\text{count}(0 < MGD(x, y))} \sum_{y=0}^{h-1} \sum_{x=0}^{w-1} MGD(x, y) \quad (6)$$

식 (6)에서 T_b 는 MGD영상을 이진영상으로 변환시킬

때 기준이 되는 임계값이고, w, h 는 컬러영상 전체의 가로와 세로길이를 의미한다.

(a) MGD 영상

(b) 이진화된 MGD 영상

그림 4. 밝기 변화량을 이용한 MGD 및 이진화 영상

4. 후처리

후처리 단계는 이진화된 MGD 영상에서 노이즈 제거와 영역 병합, 문자 블록 설정 과정을 포함한다. 이진화된 MGD 영상은 대체로 문자영역에서 MGD값이 존재하고 비문자 영역에서는 남아있지 않지만, 그림과 배경과의 경계부분이나 영상 내 급격한 밝기 변화값을 가지는 지점에서 남아있는 경우가 있다. 또한 문서 영상의 스캔 과정에서 발생할 수 있는 잡음 때문에 의도하지 않게 MGD값이 남을 수도 있다. 따라서 이들을 제거할 수 있도록 줄 단위로 MGD 길이를 구하여 $2n-1$ 보다 작은 경우 이를 제거한다(n 은 문자의 폭, 식 (4) 참조). 이는 문자가 대체로 단독으로 존재하지 않고 문자열로써 존재하는 특성을 이용하는 것이다. 다음으로 산재되어 있는 각각의 MGD를 병합시키기 위해 팽창(dilate)과 침식(erode)의 모폴로지(morphology) 연산을 수행하는 영역 병합 과정을 거친다. 이때 모폴로지 연산의 필터 크기는 식 (5)의 n 과 동일한 값으로 설정하였다. 마지막으로 병합된 MGD 좌표들의 시작점과 끝점을 이용하여 문자 블록을 설정한다. [그림 5]는 [그림 4]의 (b)에 대한 후처리 단계의 결과를 나타낸다.

그림 5. 후처리 결과 영상

5. 문자 레이블 결정

문자 레이블 결정 단계는 후처리 후 설정된 문자 블록과 레이블 각각의 화소 위치를 고려하여 각 레이블의 속성을 문자 또는 비문자로 분류하는 과정이다. 즉, 앞선 레이블링 단계에서 추출된 각 레이블이 [그림 5]와 같이 나타난 문자 블록의 범위 내 존재하는지 확인하여 해당 레이블을 문자 또는 비문자로 분류한다. [그림 6]은 문자 레이블 결정 단계의 pseudo code를 나타낸다.

```

/* 레이블의 문자 또는 비문자 분류 */

label[NL]; // 레이블 배열
// NL은 레이블의 총 수
for (i=0; i < NL; i++) {
    label[i].text = false;
    decLblText(label, i);
}

function decLblText(label, i) {
    for (y=label[i].start_y; y < label[i].end_y; y++)
        for (x=label[i].start_x; x < label[i].end_x; x++)
            if (label[i].position is in text block) {
                label[i].text = true;
                return;
            }
}

```

그림 6. 문자 레이블을 결정하는 pseudo code

III. 실험 및 평가

본 논문에서 제안한 레이블링 기법과 밝기값 변화에 기반한 문자영역 추출 방법의 성능평가를 위하여 OpenCV를 이용한 C 기반의 프로그램을 구현하였으며 [18], 다양한 컬러영상으로 실험을 수행하였다. 실험에서는 먼저 기존 방법과 제안한 방법으로 추출된 결과영상을 비교 및 분석하였고, 다음으로 정확도 및 재현율

과 비문자 추출의 역 비율을 측정하여 기존 방법과 제안한 방법의 성능을 비교하였다.

1. 결과 영상 분석

제안한 방법의 결과를 분석하기 위해 [그림 7]과 같이 논문, 교과서, 웹, 비디오 영상 등 4가지 형태의 문자가 포함된 영상에 대하여 원본영상과 기존방법, 제안한 방법을 비교하였다. [그림 7]의 위에서부터 아래까지는 각각 논문, 교과서, 웹, 비디오 영상을 나타낸다. 기존 방법의 비교대상으로는 (b) 밝기값 변화를 이용한 방법 [3]과 (c) 미디언 필터링 기반의 방법[4, 5]을 선택하였다. [그림 7]에 사용된 실험영상은 모두 컬러영상으로 해상도는 논문 영상에서부터 비디오 영상까지 각각 794x932, 878x1220, 352x651, 640x272 이다. [그림 7]의 논문 영상에 대하여 Shivakurama[3] 방법을 이용한 (b)의 경우 좌측 상단의 바운딩 박스의 일부분과 좌측 중단의 그림들을 문자영역으로 잘못 인식하였다. 이 방법은 밝기값 변화만을 이용하여 문자영역을 추출하기 때문에 결과에 나타난 것처럼 문자와 근접한 바운딩 박스의 일부가 문자열로 인식되는 문제점이 있다. 또한 밝기값 변화가 큰 그림의 경우 문자영역으로 잘못 인식하는 문제점을 보이고 있다. 논문 영상에 대하여 Jang[4, 5]의 방법을 이용한 (c)의 경우는 좌상단의 바운딩 박스와 우상단의 표가 문자로 잘못 인식되었다. 이 방법은 문자 개체와 인접하게 위치한 개체를 문자 개체로 분류하는 과정이 있고, ‘—’와 ‘|’ 같은 모음을 문자로 분류하기 위하여 문자의 최대높이를 기준으로 그 이하의 높이인 개체의 경우 문자로 처리하는 과정이 있다. 이 과정들로 인해 바운딩 박스가 개체의 흑화소 밀도 측정으로 먼저 비문자로 분류가 되었으나 바운딩 박스 내부의 문자열과 인접하게 위치하기 때문에 문자로 인식되었고, 직선으로 이루어진 표를 먼저 비문자 개체로 분류하였으나 선의 높이가 문자의 최대높이보다 작아 문자로 인식되는 문제점이 있었다. 논문 영상에 대하여 제안한 방법을 이용한 (d)의 경우 결과영상의 좌중단 화살표 그림이 문자로 인식되었다. 이는 문자가 화살표 그림에 속해 있어서 문자와 화살표 그림이 동일한 레이블로 설정되어 화살표 그림까지 함께 문자로 인식되었

그림 7. 기존방법과 제안한 방법의 문자영역 추출 결과

다. 그러나 그 외 비문자 영역은 모두 제거되어 문자영역이 매우 정확하게 추출되었음을 확인할 수 있었으며, 교과서, 웹 영상의 경우 역시 다른 기존의 방법들에 비하여 비교적 문자 추출 결과가 우수하였다. 그러나 비디오 영상의 경우 Shivakurama[3] 방법은 정확하게 문자를 추출한데 비하여 제안한 방법은 정확하게 추출하지 못하였다. 비디오 영상은 논문, 교과서, 웹 유형의 문

서영상과 달리 비문자 영역에 문자영역이 포함되기 때문에 제안한 방법의 레이블링 또는 레이블 필터링 과정에서 문자가 손실되는 문제점을 보였다. 이러한 문제점을 해결하기 위해서는 비디오 영상과 문서 영상을 분류할 수 있는 방법을 추가하거나, 비디오 영상의 경우 레이블링 및 레이블 필터링 과정을 거치지 않도록 하며, 또는 반전된 영상으로 처리할 수 있는 알고리즘을 추가

하여 성능을 높일 필요가 있다.

2. 성능 비교

제안한 방법과 기존 방법의 수치적인 성능 비교를 위하여 정확도(precision)와 재현율(recall)을 측정하였다. 측정에 필요한 기준은 화소, 문자, 문자열 등 여러 단위가 있지만[3][13][19][20], 본 논문은 문자 단위로 측정하였으며, 식 (7)을 이용하였다.

$$\begin{aligned} Precision &= \frac{TNOCT}{NODT} 100 \\ Recall &= \frac{TNOCT}{TNOT} 100 \end{aligned} \quad (7)$$

여기서 TNOCT는 추출된 문자 중 정확한 문자개체의 수(total number of correct texts), NODT는 추출된 개체의 수(number of detected texts), TNOT는 실제 문자개체의 총 수(total number of texts)를 의미한다.

[표 1]은 기존의 밝기값 변화를 이용하여 문자를 추출하는 방법[1-3]과 연결요소와 메디안 필터링 기반의 방법[4][5], 제안하는 방법의 성능을 나타내고 있다. 실험영상으로는 문자와 비문자가 적절히 혼합된 논문, 교과서, 웹 사이트, 비디오 영상 등의 컬러 영상을 사용하였다. 논문은 다단으로 작성되어 있고, 한글논문 2개, 영어논문 3개를 선택하여 한페이지 단위로 총 5장을 사용하였다. 교과서는 그림이 많은 고교 수학 및 과학 과목을 한페이지 단위로 선택하였으며 한글과 영어, 기호, 숫자가 혼합된 5장의 영상을 이용하였다. 웹 사이트는 일반적으로 접할 수 있는 국내, 국외 기사와 홈페이지 등의 10개 주소를 선택하여 문자와 비문자가 혼합될 수 있도록 화면을 캡처하여 10장의 영상을 사용하였으며, 비디오 영상은 4개의 한글 자막이 포함된 영상과 1개의 영어 자막이 포함된 영상을 사용하였다. 그리고 각 실험영상의 유형에 따라 문자 개체는 9369, 2572, 12212, 62개가 포함되었으며, 비문자 개체는 각각 395, 131, 113, 5개가 포함되었다. [표 1]의 정확도 및 재현율의 평균(average)은 실험영상의 유형과 관계없이 모든 문자에 대한 정확도 및 재현율을 의미한다. [표 1]에서 나타난 바와 같이 제안한 방법과 기존의 방법들 모두 높은

정확도와 재현율이 측정되었으며, 격차 또한 크지 않다. 이는 [그림 7]의 각 방법의 결과 (b), (c), (d)와 같이 실제 시각적인 차이가 있음에도 불구하고 문자의 개체 수가 비문자의 개체 수보다 훨씬 커서 수치적인 정확도 및 재현율에 영향을 미치지 못하기 때문이다. [그림 7]의 웹 영상을 예로 들면 문자 수가 623개, 비문자 수는 5개로써 정확도는 (b)부터 (d)까지 각각 99.11%, 99.36%, 99.84%이고, 재현율은 각각 89.41%, 100%, 100%로 측정되어 큰 차이를 나타내지 못한다. 따라서 본 논문은 식 (8)을 이용하여 비문자 추출의 역 비율([표 1]의 IRNTE)을 별도로 측정하였다.

$$IRNTE = \left(1 - \frac{NODNT}{TNONT}\right) 100 \quad (8)$$

여기서 IRNTE은 비문자 추출의 역 비율(inverse ratio of non-texts extraction, 이하 IRNTE)이며, NODNT는 추출된 비문자 개체의 수(number of detected non-texts), TNONT는 실제 비문자 개체의 총 수(total number of non-texts)를 의미한다. [표 1]에 나타난 바와 같이 제안하는 방법은 기존 방법과 비교하여 평균 IRNTE이 최소 29.82%, 최대 54.66% 높게 측정되었다. 특히 제안한 방법의 IRNTE은 비디오 영상을 제외한 논문, 교과서, 웹 사이트 등의 문서영상에서 기존의 다른 방법보다 높게 측정되었다. 이는 기존의 밝기값 변화를 이용하는 방법[1-3]이 밝기값 변화가 큰 비문자 영역이나 문자영역과 인접한 비문자 개체를 문자로 잘못 인식하는 문제점을 가지고 있고, 기존의 메디언 필터링 기반의 방법[4][5] 역시 문자와 인접한 비문자 영역을 문자로 잘못 인식하는 문제점을 가지고 있는 반면, 제안한 방법은 레이블 필터링 과정을 통해 미리 비문자 영역을 제거하여 비문자 영역이 문자영역으로 잘못 인식할 가능성을 최소화하였기 때문에 IRNTE이 높게 측정된 것으로 분석된다. 그러나 비디오 영상의 경우 제안한 방법의 재현율과 IRNTE은 기존의 밝기값 변화를 이용하는 방법보다 낮게 측정되었는데, 사진, 표, 그래프 등의 비문자 영역과 문자영역이 분리되어 배치된 문서영상과 달리 비디오 영상은 그림이나 유

색 배경과 같은 비문자 영역 내 문자가 포함되어 있기 때문이다. 그로 인해 레이블링 단계에서 문자와 그림 또는 문자와 배경이 동일한 레이블로 설정되어 문자영역이 레이블링 필터링 과정에서 제거되기 때문으로 분석된다. 따라서 레이블링 과정에서 Otsu 알고리즘[21]을 적용하거나 또는 다단계 색상 군집화 알고리즘[13][14] 등을 이용하여 레이블링 과정을 수행하면 더욱 높은 성능을 보일 것으로 예상된다.

표 1. 제안하는 방법과 기존 방법의 성능 비교

Method	Type	Precision(%)	Recall(%)	IRNTE (%)
Wong [1]	Manuscript	97.47	94.44	40.25
	School book	97.82	96.07	58.78
	Web	99.46	52.87	68.14
	Video	98.33	95.16	80.00
	Average	98.23	73.65	49.22
Kwon [2]	Manuscript	96.63	70.48	39.75
	School book	98.55	92.38	68.70
	Web	99.17	92.45	16.81
	Video	99.17	92.45	80.00
	Average	98.26	83.90	41.93
Shivakumara [3]	Manuscript	96.92	96.06	23.04
	School book	97.02	79.82	50.38
	Web	99.14	91.01	15.04
	Video	95.75	72.58	80.00
	Average	98.03	91.73	27.64
Jang [4, 5]	Manuscript	97.93	96.99	51.39
	School book	98.33	96.07	69.95
	Web	99.38	99.01	37.17
	Video	95.16	95.16	40.00
	Average	98.70	97.91	52.48
제안한 방법	Manuscript	99.31	97.64	80.00
	School book	99.92	99.26	96.95
	Web	99.75	99.48	74.34
	Video	97.50	62.90	60.00
	Average	99.59	98.65	82.30

3. 잡음 제거 분석

잡음이 있는 영상에 대한 결과영상의 분석을 위해 실제 문서영상을 스캔하여 제안한 방법을 적용하였다. [그림 8]의 (a) 스캔 된 원본 영상은 794x932 해상도로써 전반적으로 스캔으로 인한 퍼짐현상이 발생하였고, 좌상단에 문서의 일부가 접혀 있다. 또한 가장자리에

문서의 테두리로 인하여 경계선이 표시되어있으며, 곳곳에 작은 잡음도 포함되어있다. [그림 8]의 (b)와 같이 제안한 방법을 적용하였을 때 문서의 접힌 부분과 잡음이 제거되었고, (a)의 좌하단과 같이 점선 형태의 테두리는 제거하지 못하였으나 이외의 경계선은 모두 제거되었다. 그러나 스캔으로 인한 퍼짐현상 때문에 (b)의 우상단에 표와 겹쳐진 'g'와 'p', 괄호 등은 표와 같은 레이블로 인식하여 함께 제거되었으며 이와 가까운 위치의 다른 문자들도 퍼짐현상으로 인하여 함께 제거되는 문제점을 보였다.

그림 8. 잡음이 있는 영상에 대한 문자추출 결과

IV. 결론

본 논문은 레이블링 기법과 밝기값 변화에 기반한 컬러영상의 문자영역 추출 방법을 제안하였다. 레이블링 및 레이블 필터링을 이용하여 비문자 영역을 미리 제거할 수 있어 비문자 영역의 잘못된 추출 가능성을 최소화 하였고, 문자영역에서 화소 단위의 밝기값 차이 특성을 이용하여 문자영역을 효과적으로 추출하였다. 실험결과 기존 방법들과 비교하여 문자 추출에 대한 정확도 및 재현율이 상당히 높았고, IRNTE 역시 높게 측정되었다. 향후 비디오 영상과 같이 유색문자와 유색배경의 혼합된 컬러영상의 정확한 문자영역 추출을 위해 알고리즘을 보완할 것이며, 점선에 대한 필터링 처리, 기울어진 영상에 대한 보정, 사선 또는 세로방향의 문자열도 인식할 수 있는 방법 등을 추가할 것이다. 또한 광학문자인식 시스템과의 연계를 통해 최종적으로 복잡

한 컬러영상에 대한 정확한 문자인식이 가능한 연구도 필요할 것으로 보인다.

참 고 문 헌

- [1] E. K. Wong and Minya Chen, "A new robust algorithm for video text extraction," Pattern Recognition, Vol.36, No.6, pp.1397-1406, 2003.
- [2] 권재현, 손창환, 박태용, 조양호, 하영호, "문자 영역을 강조하기 위한 적응적 오차 확산법", 대한전자공학회 논문지, 제43권, 제1호, pp.9-17, 2006.
- [3] Palaiahnakote Shivakumara, Trung Quy Phan, and Chew Lim Tan, "A Gradient Difference Based Technique for Video Text Detection," Proceedings of International Conference on Document Analysis and Recognition, pp.156-160, 2009.
- [4] 장대근, 황찬식, "이미지 필터와 제한조건을 이용한 문서영상 구조분석", 정보처리학회 논문지B, 제9권, 제3호, pp.311-318, 2002.
- [5] 장대근, 황찬식, "연결요소 방법과 메디안 필터를 이용한 문서영상 기하학적 구조분석", 한국통신학회 논문지, 제27권, 제8호, pp.805-813, 2002.
- [6] C. Strouthopoulos, N. Papamarkos, and A. E. Atsalakis, "Text extraction in complex color document," Pattern Recognition, Vol.35, No.8, pp.1743-1758, 2002.
- [7] B. T. Chun, Y. L. Bae, and T. Y. Kim, "Text extraction in videos using topographical feature of characters," IEEE International Conference on Fuzzy System, Vol.2, pp.1126-1130, 2002.
- [8] 전병태, 송치양, "주파수 특성과 역전파 신경망 알고리즘을 이용한 문자 영역 분할 방법", 한국컴퓨터정보학회 논문지, 제11권, 제4호, pp.55-60, 2006.
- [9] 이형렬, "복잡한 컬러 문서에서의 문자추출 및 문자인식", 김천과학대학 논문집, 제27권, pp.111-115, 2001.
- [10] Sachin Grover, Kushal Arora, and Suman K. Mitra, "Text Extraction from Document Images using Edge Information," IEEE India Council Conference, pp.1-4, 2010.
- [11] Q. Yuan and C. L. Tan, "Text Extraction from Gray Scale Document Images using Edge Information," Proceedings of International Conference on Document Analysis and Recognition, pp.302-306, 2002.
- [12] 박종천, 황동국, 이우람, 전병민, "적응적 문자-에지 맵을 이용한 다양한 기울기와 크기를 갖는 텍스트 영역 검출", 한국콘텐츠학회 종합학술대회 논문집, 제5권, 제1호, pp.5-9, 2007.
- [13] Y. J. Song, "Text Region Extraction and Text Segmentation on Camera-captured Document Style Images," Proceedings of International Conference on Document Analysis and Recognition, Vol.1, pp.172-176, 2005.
- [14] 김지수, 김수형, "자연영상 텍스트 이진화를 위한 3단계 색상 군집화 알고리즘", 정보처리학회 논문지B, 제12권, 제7호, pp.1-8, 2005.
- [15] 김의정, "칼라 문서에서 문자 영역 추출 및 문자 분리", 한국 퍼지 및 지능시스템학회 논문지, 제9권, 제4호, pp.444-450, 1999.
- [16] K. C. Jung and J. H. Han, "Hybrid approach to efficient text extraction in complex color images," Pattern Recognition, Vol.25, No.6, pp.679-699, 2004.
- [17] 정기철, "복잡한 영상 내의 문자영역 추출을 위한 텍스처와 연결성분 방법의 결합", 대한전자공학회 논문지, 제41권, 제6호, pp.175-186, 2004.
- [18] G. Bradski and A. Kaehler, *Learning OpenCV - Computer Vision with the OpenCV Library*, O'Reilly Media, 2008.
- [19] Sameer Antani, Ullas Gargi, David Crandall, Tarak Gandhi, and Rangachar Kasturi, "Extraction of Text in Video," Technical Report, CSE-99-016, 1999.

- [20] A. Wernicke and R. Lienhart, "On the segmentation of text in videos," IEEE International Conference on Multimedia and Expo, Vol.3, pp.1511-1514, 2000
- [21] Nobuyuki Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, Man and Cybernetics, Vol.9, No.1, pp.62-66, 1979.

- 2001년 ~ 2002년 : EG Technology Inc. Software Engineer
- 2004년 ~ 2006년 : 삼성전자 책임연구원
- 2006년 ~ 현재 : 경원대학교 컴퓨터공학과 조교수
<관심분야> : 영상처리, 신호처리, 3D 그래픽스, 멀티미디어

저 자 소 개

원 종 길(Jong-Kil Won)

준회원

- 2010년 2월 : 경원대학교 멀티미디어학과(공학사)
- 2010년 3월 ~ 현재 : 경원대학교 전자계산학과 석사과정

<관심분야> : 영상처리, 멀티미디어

김 혜 영(Hye-Young Kim)

준회원

- 2011년 2월 : 경원대학교 컴퓨터미디어학과(공학사)
- 2011년 3월 ~ 현재 : 경원대학교 전자계산학과 석사과정

<관심분야> : 영상처리, 멀티미디어

조 진 수(Jin-Soo Cho)

정회원

- 1994년 : 인하대학교 전자공학과(공학사)
- 1998년 : Columbia University 전자공학과(공학석사)
- 2003년 : Georgia Institute of Technology 전자공학과(공학박사)

사)