스크린 이미지 부호화를 위한 에지 정보 기반의 효과적인 형태학적 레이어 분할

Effective Morphological Layer Segmentation Based on Edge Information for Screen Image Coding

박상호, 이시웅

한밭대학교 정보통신전문대학원

Sang-Hyo Park(ptagmouse@nate.com), Si-Woong Lee(swlee69@hanbat.ac.kr)

요약

다중 레이어 영상 모델인 Mixed Raster Content 모델 (MRC) 기반의 영상 부호화는 스크린 이미지와 같은 혼합 영상을 전경 레이어, 이진 마스크 레이어, 배경 레이어로 재구성한 뒤, 각 레이어마다 그 레이어의 신호 특성에 적합한 부호화기를 이용하여 영상을 압축하는 기법이다. 문자와 같은 계단 형태의 강한 에지를 갖는 영역의 위치 정보를 마스크 레이어에 저장하고, 그 위치의 색상 신호는 전경 레이어에 저장한다. 그리고 나머지 영역인 배경 영역의 색상 신호는 배경 레이어에 저장한다. 따라서 마스크 레이어가 전경과 배경의 분할 정보를 담게 되며, 이 분할 정보의 정확도에 따라 전체 부호화기의 압축 효율이 직접적인 영향을 받는다. 본 논문은 MRC 기반의 영상 부호화를 위한 새로운 레이어 분할 알고리즘을 제안한다. 제안 방법은 형태학적 필터링 top hat 변환을 이용하여 문자를 배경신호로부터 분할한다. 이때 문자의 경계를 에지 맵으로부터 추정하여 문자 색상과 배경과의 상대적 밝기를 결정하고 이를 통해 형태학적 필터링에 필요한 top hat 변환의 종류를 정확히 선택하도록 하였다. 실험을 통해 제안 방법이 비교 대상 알고리즘에 비해 우수한 분할 성능을 가짐을 보인다.

■ 중심어 : | 스크린 이미지 부호화 | MRC 모델 | 형태학적 레이어 분할 |

Abstract

An image coding based on MRC model, a kind of multi-layer image model, first segments a screen image into foreground, mask, and background layers, and then compresses each layer using a codec that is suitable to the layer. The mask layer defines the position of foreground regions such as textual and graphical contents. The colour signal of the foreground (background) region is saved in the foreground (background) layer. The mask layer which contains the segmentation result of foreground and background regions is of importance since its accuracy directly affects the overall coding performance of the codec. This paper proposes a new layer segmentation algorithm for the MRC based image coding. The proposed method extracts text pixels from the background using morphological top hat filtering. The application of white or black top hat transformation to local blocks is controlled by the information of relative brightness of text compared to the background. In the proposed method, the boundary information of text that is extracted from the edge map of the block is used for the robust decision on the relative brightness of text. Simulation results show that the proposed method is superior to the conventional methods.

■ keyword : | Screen Image Coding | MRC Model | Morphological Layer Segmentation |

* 이 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임.(No. 2012R1A1A4A01014227)

접수일자 : 2013년 09월 05일

수정일자 : 2013년 10월 30일

심사완료일 : 2013년 10월 22일

교신저자 : 이시웅, e-mail : swlee69@hanbat.ac.kr

1. 서론

카메라에 의해 촬영되는 자연 영상과 달리 웹 페이지, PDF 파일, 슬라이드, 컴퓨터 화면 등과 같이 컴퓨터 또는 스마트폰 등의 다양한 전자기기에 의해 생성된 영상을 스크린 이미지라 한다. 스크린 이미지는 일반적으로 문자/그래픽 영역과 자연영상 영역이 같이 혼재되어 있으며, 이 중에서 문자/그래픽 영역은 자연영상 영역과는 구별되는 통계적 특성들을 갖고 있다. 이러한 강한 이방성을 갖는 스크린 이미지의 특성상 자연영상의 효율적인 압축을 목적으로 개발된 기존의 표준 부호화기를 스크린 이미지에 그대로 적용하는 것은 압축 성능 면에서 매우 비효율적이다. 따라서 이를 보완하고 스크린 이미지의 특성을 잘 고려함으로써 고효율의 압축이 가능한 새로운 부호화 기술에 대한 연구가 활발히 진행되고 있다.

최근에 스크린 이미지와 같은 혼합 영상을 위한 새로운 부호화 방안으로 레이어 기반의 Mixed Raster Content (MRC) model[1]이 제안되었다. MRC 모델은 입력 영상을 [그림 1-(a)]와 같이 전경 레이어, 마스크 레이어, 그리고 배경 레이어로 분할한다. 마스크 레이어는 입력 이미지에서 문자 영역에 속하는 픽셀들의 위치 정보를 나타내고, 전경 레이어는 이진 마스크 레이어에 기록된 픽셀들, 즉 문자 영역의 색상 정보를 저장한다. 이 두 레이어는 함께 하나의 쌍을 구성하며 [그림 1-(b)]와 같이 동일한 배경 레이어 위에 다수의 전경/마스크 쌍이 형성될 수도 있다. 배경 레이어는 전경 영역을 제외한 나머지 영역들에 대한 색상 정보를 저장한다. 이와 같이 MRC 모델의 부호화 방식은 입력 이미지를 동질의 특성을 갖는 3계층의 레이어로 분할한 뒤, 각 레이어 별로 그 레이어의 신호 특성에 적합한 방식으로 독립적 부호화를 수행함으로써 부호화 효율을 최적화하는 압축 기법이다.

MRC 모델에서 가장 중요한 레이어는 배경 영역과 문자 영역의 위치 정보를 담고 있는 마스크 레이어라고 할 수 있다. 이는 입력 영상을 문자의 전경 영역과 자연영상의 배경 영역으로 구분하는 분할의 정확도에 따라 전체 부호화기의 압축 효율이 결정되기 때문이다. 입력

영상을 레이어로 분할하기 위한 기존의 기법들은 크게 밝기 기반 분할 방식, 에지 기반 분할 방식 및 형태학적 분할 방식으로 분류할 수 있다.

(a) 일반적인 MRC 모델

(b) 다중 쌍(전경/마스크)으로 구성된 MRC 모델

그림 1. MRC 구조

밝기 기반의 분할 방식은 문자 영역이 배경에 비해 상대적으로 밝거나 어둡다는 가정을 바탕으로 입력 영상의 밝기를 임계치화 함으로써 전경과 배경을 분할하는 방식이다[2]. 배경과 전경 영역이 확연히 구분되는 단순한 영상에 대해서는 단일 임계치를 사용하여 간단하면서도 우수한 분할 결과를 나타낼 수 있다. 그러나 문자 영역의 색상이 다양하거나 복잡한 배경을 갖는 영상에 대해서는 분할의 결과가 매우 부정확하다. 이를 개선하기 위해 Niblack[3]와 Sauvola[4]는 지역 임계값 분할 방식을 이용하였다. 그리고 Y. Chen[5]은 지역적으로 최적의 임계값을 이용하여 클러스터링을 하는 방식으로 영상을 분할하는 방식을 제안하였다. 하지만 이러한 밝기 정보만을 이용한 분할 방식은 앞서 언급한 문제에서 여전히 자유로울 수가 없으며, 다양한 스크린 이미지 상에 존재하는 문자요소를 추출하는데 있어 한

계가 있다.

예지 기반의 분할 방식은 문자의 경계에서 강한 예지가 형성되는 특성을 바탕으로 영상의 예지 정보를 이용하여 문자를 추출하는 방식이다. Y. Liu[6]와 C.Jung[7]은 외곽선이 생성된 문자의 stroke 속성을 이용하는 방식을 제안하고 있다. Stroke를 이용한 문자 추출은 입력 이미지의 종류가 국한 된다는 단점이 있으며 문자 이외의 영역에 대한 후처리 문제를 수반한다. 박종천[8]은 폐쇄된 예지 정보를 이용하여 문자 후보 영역을 추출하는 방식을 제안하였다. 이 방식은 추출된 예지가 폐쇄되어 있지 않을 경우 문자 추출의 정확도가 낮아진다는 단점을 갖는다. 원종길[9]은 레이블 필터링을 통한 문자 영역을 추출하는 방식을 제안하였다. 이 방식은 레이블의 크기 정보를 이용하여 문자 후보 영역을 추출하고 있으나 다양한 크기의 문자와 복잡한 배경 신호를 갖는 스크린 이미지에서의 문자 추출에 적합한 방식이라고 볼 수 없다. Xin Zhang[10]은 Wonjun Kim[11]이 제안한 트랜지션 맵과 예지 정보를 이용하여 문자 영역을 추출하는 기법을 제안하였다. 트랜지션 맵 모델은 밝기의 예지뿐만 아니라 채도의 예지까지 고려하여 문자 영역을 추출하는 방식이다. 그러나 이 방식은 트랜지션 맵과 예지 맵을 통합할 때 적용되는 가중치에 따라 추출의 정확도가 달라질 수 있다는 단점을 갖는다.

형태학 기반의 분할 방식은 형태학적 필터를 사용하여 문자와 배경 영역의 상대적인 밝기차 및 구조요소의 크기와 형태를 기반으로 문자 영역을 추출하는 방식이다. 즉 구조요소의 크기와 형태, 그리고 형태학적 연산 순서 등을 활용하여 분할하고자 하는 문자 요소의 형태학적 특성을 기반으로 영역 분할을 수행함으로써 임계치가 필요 없는 효과적인 영역 분할이 가능하다는 장점을 갖는다. 이러한 장점에도 불구하고 형태학적 분할 방식은 적절한 형태학적 필터 선택의 부정확성으로 인해 전체적인 분할 성능이 제한된다는 문제점을 갖고 있다. 장인영[12]은 1x5 구조요소로 열기와 닫기 연산을 수행하여 문자 영역을 추출하고 있다. 이 방식은 수평 방향의 문자만을 고려하고 있으며 문자의 정확한 영역을 추출할 수 없다는 단점이 있다. 정확한 문자 영역을 추출하기 위해 Juliet[13]은 top hat 연산을 이용하여 문

자를 추출하는 방식을 제안하였다. Juliet은 white top hat과 black top hat 연산을 적용하기 위해 입력 영상을 Otsu 임계치를 이용한 전역 분할 방식으로 사전 분할한 결과를 이용하였다. 즉, 입력 영상을 밝기가 대비되는 두 영역으로 분할한 후 밝은 영역에 white top hat 연산을 적용하고 어두운 영역에 black top hat 연산을 적용하였다. 그러나 이 방식은 밝은(어두운) 영역에서는 문자가 배경 신호에 비해 상대적으로 어둡다(밝다)는 가정을 기반으로 설계된 것이다. 따라서 이 가정이 만족되기 어려운 다양한 문자 색상을 갖는 복잡한 영상에서는 분할의 성능이 심각하게 저하될 수 있다. 뿐만 아니라, 전역적 방식을 이용한 밝기 분할 방식은 문자 색상의 다양성과 그림자와 같은 문자 속성 등에 의해 추출된 문자의 불연속과 노이즈가 발생하여 문자 추출의 정확도가 저하될 수 있다는 문제점이 존재한다.

본 논문에서는 기존의 이러한 문제점을 해결하기 위해 예지 정보를 활용한 새로운 형태학적 레이어 분할 알고리즘을 제안한다. 제안 방식에서는 문자 영역의 경계를 예지 정보로부터 추정하여 문자 주변 영역의 밝기 정도를 계산하고 이를 기반으로 문자 영역이 배경에 비해 상대적으로 밝은지에 대한 판정을 국부적으로 수행하도록 설계하였다. 문자와 배경간의 강한 예지에서만 예지가 추출되도록 예지 검출기를 구성함으로써 제안 알고리즘의 정확도를 확보할 수 있었으며, 국부적인 top hat 연산의 선택을 통해 복잡한 색상의 문자 영역도 효과적으로 추출 가능하도록 하였다. 실험을 통해 제안 방식이 다양한 스크린 영상에 대해 기존의 방식보다 매우 정확한 레이어 분할 성능을 가짐을 확인하였다.

본 논문의 구성은 다음과 같다. 2장은 형태학적 레이어 분할에 대한 기본 개념에 대하여 서술한다. 3장은 제안 방식에 대하여 서술하고 4장은 실험을 통해 제안 방식의 성능을 확인한다. 마지막으로 5장에서 결론을 맺는다.

2. 형태학적 레이어 분할의 개요

형태학적 레이어 분할은 형태학적 필터의 구조 요소에 포함된 문자 요소의 형태와 문자와 배경 영역의 상대적인 밝기 차이를 이용하여 입력 영상을 전경과 배경

영역으로 분할하는 방식을 말한다. 모든 형태학적 연산 및 필터는 아래 식(1)로 정의되는 침식과 팽창이라는 두 개의 기본 연산의 조합으로 구성된다.

$$(f \ominus B)(x, y) = \min_{(x', y' \in B)} f(x + x', y + y') \quad (1)$$

$$(f \oplus B)(x, y) = \max_{(x', y' \in B)} f(x + x', y + y')$$

여기에서 f 와 B 는 각각 입력 영상과 형태학적 연산을 위한 구조요소(structuring element)를 나타낸다. 침식 연산은 구조요소에 포함된 원 신호 중 최소값을 출력으로 하는 연산이며 팽창 연산은 이와 반대로 최대값을 출력으로 하는 연산이다.

이 기본 연산으로부터 형태학적 열기 연산과 닫기 연산을 다음 식(2)와 같이 정의한다.

$$\begin{aligned} f \circ B &= (f \ominus B) \oplus B \\ f \bullet B &= (f \oplus B) \ominus B \end{aligned} \quad (2)$$

열기 연산은 입력 영상을 구조요소를 이용하여 침식 후 팽창 연산을 수행하는 것이다. 따라서 열기 연산은 입력 영상에서 주변 영역보다 밝은 구조 요소보다 작은 영역들을 제거하고, 이와 반대로 닫기 연산은 주변 영역보다 어두운 작은 영역들을 제거한다.

형태학적 레이어 분할에서 사용되는 top hat 연산은 열기와 닫기 연산을 이용하여 다음 식과 같이 정의된다.

$$\begin{aligned} T_W(f) &= f - f \circ B \\ T_B(f) &= f \bullet B - f \end{aligned} \quad (3)$$

$T_W(f)$ 는 white top hat 연산이며 $T_B(f)$ 는 black top hat 연산이다. white top hat 연산은 원본 영상과 열기 연산으로 구해진 결과 영상과의 차 영상을 구하여 열기 연산에 의해 사라진 밝은 영역만 남기게 된다. 이와 반대로 black top hat 연산은 닫기 연산의 결과 영상과 원본 영상의 차 영상을 구하게 되어 닫기 연산에 의해 사라진 어두운 영역만을 남기게 된다. 따라서 문자 영역이 배경에 비해 상대적으로 밝은 경우에는 white top hat 연산을, 반대의 경우에는 black top hat 연산을 적용함으로써 문자를 배경으로부터 추출할 수 있다. 식(3)

의 top hat 연산 중 어느 것이 현재의 문자 추출에 적합한 연산인지를 정확히 결정하는 것은 정확한 레이어 분할을 위한 필수 조건이며, 이는 문자 영역이 배경에 비해 상대적으로 밝은지 어두운지로부터 결정될 수 있다. 문자 영역과 배경이 다양한 색상을 갖는 복잡한 영상일수록 이러한 결정은 전역적(global) 방식으로는 수행될 수 없으며, 동일한 밝기 특성의 국부(local) 영역별로 결정되어야 한다.

3. 제안 방법

그림 2. 제안 방법의 흐름도

[그림 2]는 제안 방식의 전체 흐름도를 보여준다. 그림에서 보듯이 제안 방법은 문자 영역을 추출하기 위해 형태학적 연산인 top hat 연산을 이용한다. 우선 입력 영상을 그레이 이미지로 변환한 후 위치에 따라 색상이 가변적인 문자들을 추출하기 위해 영상을 특정 크기의 블록으로 분할하여 블록 단위의 국부적인 문자 추출 방식을 적용한다. 즉, 각 분할된 블록마다 문자 영역의 상대적인 밝기 정보를 추정하고 이를 기반으로 그 블록에 적합한 top hat 연산의 종류를 결정하여 형태학적 문자 추출을 수행한다.

분할된 각 블록 영역 내 문자의 밝기를 조사하기 위해 각 블록의 에지 맵을 생성한다. 이 때 에지는 배경 신호의 경사에서 추출되는 것이 아니라 문자와 배경 간

의 계단형 경계에서만 추출되도록 하기 위해 약한 에지를 제외한 강한 에지만을 추출하도록 설계하였다. 이는 배경 영역 대비 문자 영역의 밝기 정도를 판단하기 위하여 에지 픽셀을 중심으로 생성하는 3x3 마스크 내부에 문자 영역을 반드시 포함시키기 위함이다. 강한 에지만을 생성하기 위하여 에지 추출 알고리즘으로 Canny Edge Detector[14]를 적용하였다. Canny 알고리즘은 기울기가 가장 큰 에지 픽셀만을 남겨 다른 에지 검출기보다 단일 에지 응답이 강해 강한 에지만을 추출할 수 있다는 장점이 있다.

배경 영역 대비 문자 영역의 밝기 정도를 판단하기 위하여 에지 픽셀을 중심으로 3x3 크기의 마스크를 생성한 후 마스크 내의 평균 밝기를 조사한다. 그림 3에서 볼 수 있듯이 마스크 내에서 문자 영역은 약 1/3 또는 2/3의 영역을 차지하며 평균적으로 마스크 영역의 약 50%를 차지한다. 마스크의 평균 밝기와 블록의 평균 밝기를 비교할 때, 블록 내의 문자 영역 비율이 에지에서 생성한 마스크 내의 문자 영역 비율보다 낮다면 마스크의 평균 밝기는 블록의 평균 밝기보다 문자의 밝기에 더 가까운 값을 갖게 된다.

그림 3. 마스크 연산을 이용한 문자 영역의 밝기 판별

표 1. 문서의 배경과 문자 픽셀 비율 통계

Font	Size	줄간격	장평	자간	Bold (Y/N)	문자비율	서원스별 문자비율 평균(%)	전체 평균(%)
Arial black	10	160	100	0%	N	11.53	12.44	8.74
Arial black	10	160	100	-5%	N	12.53		
Arial black	10	100	95	0%	N	10.73		

Hand writing	10	160	100	0%	N	2.97	8.50	
Hand writing	10	160	100	0%	Y	9.81		
Hand writi	12	160	100	-5%	Y	9.95		
Hand writi	14	100	100	-5%	N	8.29		

Tahoma	10	160	100	0%	N	2.39	8.98	
Tahoma	10	160	100	0%	Y	8.67		
Tahoma	12	160	100	0%	Y	9.99		
Tahoma	14	160	100	-5%	Y	14.27		

Times roman	10	160	100	0%	N	3.19	5.12	
Times roman	10	160	100	0%	Y	5.20		
Times roman	12	100	100	0%	N	6.36		
Times roman	14	160	100	0%	Y	7.69		

[표 1]은 문서 내에 포함된 문자와 배경간의 비율에 대하여 조사한 것으로 문서에서 배경과 문자 영역의 통계적인 비율을 보여 준다. 동일한 문자열의 폰트와 크기, 줄 간격 그리고 장평과 자간 등에 변화를 주어 문자 영역의 비율을 조사한 결과 배경 영역 대비 문자 영역이 최대 15%를 넘지 않고 평균적으로 10% 미만의 비율을 보인다. 따라서 블록이 문서의 통계적인 비율을 따른다면 에지에서 생성한 마스크의 문자 비율(50%)이 블록의 문자 비율보다 높으며, 이는 에지 마스크의 평균 밝기와 블록의 평균 밝기 비교에 의한 문자의 밝기 판단이 문자 추출의 적합한 근거가 된다는 것을 의미한다.

블록이 문서의 통계적 비율을 따르도록 하기 위하여 블록은 형태학적 연산에 사용되는 구조 요소의 두 배 이상의 크기로 설정한다. 구조 요소에 하나의 문자 요소가 포함된다고 했을 때, 블록을 문단 간의 공백이 포함 되도록 구조 요소의 두 배 이상의 크기로 설정 한다면 블록은 줄 간격과 자간 등의 속성에 의해 배경 영역의 비율이 높아져 에지에서 생성한 마스크의 문자 비율보다 낮아지게 되고 문서의 통계적 비율에 근사하게 된다. 본 논문의 실험에서는 형태학적 구조요소와 블록의 크기를 각각 5x5와 32x32로 정의 하였다.

블록 내 문자 영역의 밝고 어두운 정도를 조사하기 위해 아래 식(4)와 (5)를 이용한다.

$$\mu_{mask}(x,y) = \frac{1}{9} \sum_{j=y-1}^{y+1} \sum_{i=x-1}^{x+1} f_{i,j}(x,y) \in E_s \quad (4)$$

$$\mu_{block}(k) = \frac{1}{MN} \sum_{i=x_k}^{x_k+M} \sum_{j=y_k}^{y_k+N} f_{i,j} \quad (5)$$

식(4)에서 $\mu_{mask}(x,y)$ 는 (x,y) 위치의 에지 픽셀을 중심으로 생성된 마스크의 평균 밝기이며, E_s 는 에지 픽셀들의 집합을 나타낸다. 그리고 $f_{i,j}$ 는 (i,j) 위치에 있는 명도 영상의 밝기를 나타낸다. 식(5)의 $\mu_{block}(k)$ 는 k 번째 블록의 평균 밝기를 나타내며 M 과 N 은 블록 수평과 수직 방향의 크기이다. 블록 내부에 생성된 모든 에지 픽셀에 대하여 $\mu_{mask}(x,y)$ 를 계산하고 모든 $\mu_{mask}(x,y)$ 에 대한 평균 밝기인 μ_{mask} 를 구한 후 μ_{mask} 와 μ_{block} 을 비교하여 top hat 연산을 수행한다. 이때 각각의 블록에 white top hat이나 black top hat 연산을 선택적으로 적용하기 위한 기준으로 아래 식(6)을 사용하여 결정한다. 이 판단 기준의 정확도는 [표 2]의 FMR(False Match Rate)을 통해 확인할 수 있다.

$$TopHat = \begin{cases} White\ TopHat, & \mu_{mask} > \mu_{block} \\ Black\ TopHat, & \mu_{mask} \leq \mu_{block} \end{cases} \quad (6)$$

μ_{mask} 가 μ_{block} 보다 밝다는 것은 블록 내 문자 요소의 밝기가 블록의 평균 밝기보다 더 밝다는 것을 의미하며, 이는 블록 내부에 포함된 문자 요소의 밝기가 문자 주위의 배경 영역보다 밝다는 것을 나타낸다. μ_{mask} 가 μ_{block} 보다 어둡다는 것은 이와 반대로 블록에 포함된 문자요소의 밝기가 배경 영역보다 어둡다는 것을 의미한다. 각각의 블록에 대하여 top hat 연산을 선택적으로 적용하여 입력 영상에서 문자요소를 추출한다. 그 후 문자요소가 추출된 모든 블록의 결과를 하나의 레이어로 통합한 후 이진화를 거쳐 이진 마스크

레이어를 생성한다.

4. 실험 결과 및 분석

실험에서는 제안 방식의 성능을 검증하기 위해 Juliet[13]의 알고리즘과 AT&T에서 개발한 상용 소프트웨어 패키지인 DjVu[15]를 참조 기법으로 사용하였다. 실험 영상으로는 [그림 4]와 같이 HEVC(High Efficiency Video Coding)[16]의 class F 영상들을 사용하였으며, 레이어 분할의 객관적인 성능 평가를 위해 아래의 척도를 평가기준으로 사용하였다.

ADR(Accurate Detection Rate)

$$= \frac{\# \text{ of correctly detected text blocks}}{\# \text{ of whole text blocks}} \quad (7)$$

FMR(False Match Rate)

$$= \frac{\# \text{ of incorrectly matched text blocks}}{\# \text{ of whole text blocks}} \quad (8)$$

ADR은 전체 문자 블록 대비 주관적 손실 없이 문자 영역을 추출해낸 블록의 비율이고, FMR은 전체 문자 블록 대비 top hat 연산의 선택이 잘못된 블록의 비율이다. 전체 블록 수는 문자 요소를 포함하고 있는 블록들의 전체 개수로 정의하였다.

[그림 5]는 각각의 실험 영상에서 분할된 이진 마스크 레이어와 손실/비손실된 문자 블록에 대한 결과를 보여준다. 녹색 블록은 문자요소를 손실 없이 추출한 블록이고, 붉은색 블록(바운딩 박스로 표시)은 문자요소에서 손실이 발생한 블록이다. [그림 5-(b)]의 DjVu의 경우 블록 기반의 컬러 클러스터링으로 전경과 배경을 분리할 때, 배경 영역의 그라데이션으로 인해 인접 블록의 전경 영역과 현재 블록의 배경 영역을 같은 전경 영역으로 판단하여 일부 문자 영역을 검출하지 못했음을 확인할 수 있다. Juliet의 분할 결과인 [그림 5-(d)]는 전역 임계치 기반으로 영상을 분할한 뒤 밝은 영역과 어두운 영역에 각각 white/black top hat 연산을 적용 하였으나 영상 내 배경과 문자 색상의 다양성으로 인해 문자 영역 검출에 있어 많은 오류가 발생하였음을

확인할 수 있다. 이에 반해 제안된 방법은 [그림 5-(f)]에서 볼 수 있듯이 매우 정확히 문자 요소를 추출하였다.

[표 2]는 각 분할 방식에 대해 ADR과 FMR을 측정한 결과이다. ADR의 경우 제안 방법이 DjVu방식과 Juliet 방식에 비해 평균적으로 약 15%와 49% 정도 추출의 정확도를 개선시킴을 알 수 있다. 뿐만 아니라 FMR의 측정 결과는 제안 방식이 Juliet의 방법에 비해 top hat 연산의 결정 오류율을 큰 폭으로 감소시킴을 보여준다. 이 결과를 통해 제안 방식이 top hat 연산의 종류를 매우 정확하게 선택하고 있으며, 이를 통해 참조 기법들에 비해 문자 추출의 정확도를 큰 폭으로 개선시킨다는 것을 확인할 수 있다.

표 2. 문자 추출의 정확도 [단위 %]

	DjVu	Juliet's method		Proposed method	
	ADR	ADR	FMR	ADR	FMR
Basket Ball	70.6	0	100	100	0
China Speed	51.1	13.8	61.7	75.5	20.2
Slide Editing	89.6	64.7	39.2	94.1	1.9
Slide Show	85.9	82.2	35.1	89.2	3.3
Average	74.3	40.2	59.7	89.7	6.3

5. 결론 및 향후 연구

본 논문에서는 MRC 모델 기반으로 스크린 이미지를 효율적으로 부호화하기 위한 새로운 형태학적 레이저 분할 방법을 제안하였다. 제안 방식에서는 입력 영상의 에지 맵을 이용하여 문자 영역의 경계를 추정하고, 이 경계 영역에서의 평균 밝기와 블록의 평균 밝기와의 비교를 통해 현재 블록에 적합한 top hat 연산의 종류를 결정하였다. 실험을 통해 제안 방식이 현재 보편적으로 사용되고 있는 상용 소프트웨어인 DjVu보다 문자 검출의 정확도를 약 15% 정도 개선시킴을 확인하였다.

제안 방식에서 사용한 실험 영상은 우리가 실생활에서 자주 접하는 영상으로 자막이 입혀진 스크린 영상, 문자 요소가 들어간 게임 영상, 그리고 문자가 추가 되는 영상과 슬라이드이다. 실험 영상이 아닌 다른 영상에서도 문자의 폭에 따라 형태학적 연산의 구조 요소의 크기를 조정하여 더욱 나은 결과를 얻을 수 있다. 향후 본 알고리즘을 개선하여 다양한 문자 폭에 유연한 성능을 보여 줄 수 있도록 지속적인 연구가 필요할 것이다.

그림 4. 실험 영상

(a) 분할된 마스크 레이어-DjVu(Slide Editing)

(b) 문자 블록과 문자 손실 블록-DjVu.

(c) 분할된 마스크 레이어-Juliet's alg.

(d) 문자 블록과 문자 손실 블록-Juliet's alg.

(e) 분할된 마스크 레이어-proposed alg.

(f) 문자 블록과 문자 손실 블록-proposed alg.

(g) 분할된 마스크 레이어-DjVu(Slide Show)

(h) 문자 블록과 문자 손실 블록-DjVu.

(i) 분할된 마스크 레이어-Juliet's alg.

(j) 문자 블록과 문자 손실 블록-Juliet's alg.

(k) 분할된 마스크 레이어-proposed alg.

(l) 문자 블록과 문자 손실 블록-proposed alg.

그림 5. 분할된 이진 마스크 레이어

참고 문헌

- [1] ITU-T Recommendation T.44 Mixed Raster Content (MRC), T.44, International Telecommunication Union, 1999.
- [2] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst., Man Cybern., Vol.9, No.1, pp.62-66, 1979.
- [3] W. Niblack, *An Introduction to Digital Image Processing*, Strandberg Publishing Company Bikerood, 1985.
- [4] J. Sauvola and M. Pietaksinen, "Adaptive document image binarization," Pattern Recognit, Vol.33, No.2, pp.236-335, 2000.
- [5] Y. Chen and B. Wu, "A multi-plane approach for text segmentation of complex document images," Pattern Recognit, Vol.42, No.7, pp.1419-1444, 2009.
- [6] Y. Liu and S. Srihari, "Document image binarization based on texture features," IEEE Trans. Pattern Anal. Mach. Intell., Vol.19, No.5, pp.540-544, 1997.
- [7] C. Jung and Q. Liu, "A new approach for text segmentation using a stroke filter," Signal Process, Vol.88, No.7, pp.1907-1916, 2008.
- [8] 박종천, "적응적 문자-예지 맵을 이용한 다양한 기울기와 크기를 갖는 텍스트 영역 검출", 한국콘텐츠학회 2007 춘계종합학술대회논문집, 제5권, 제1호, pp.5-9, 2007.
- [9] 원종길, "레이블링 기법과 밝기값 변화에 기반한 컬러영상의 문자영역 추출 방법", 한국콘텐츠학회논문지, 제11권, 제12호, pp.511-521, 2011.
- [10] X. Zhang, F. Sun, and L. Gu, "A Combined Algorithm for Video Text Extraction," 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, pp.2294-2298, 2010.
- [11] W. Kim and C. Kim, "A New Approach for Overlay Text Detection and Extraction From Complex Video Scene," IEEE Trans. Image Process, Vol.18, No.2, pp.401-411, 2009.
- [12] 장인영, "형태학과 문자의 모양을 이용한 뉴스 비디오에서의 자동 문자 추출", 정보과학회논문지, 컴퓨팅의 실제, 제8권, 제4호, pp.479-488, 2002.
- [13] S. Ebenezer Juliet, V. Sadasivam, and D. Jemi Florinabel, "Effective layer-based segmentation of compound images using morphology," Journal of Real-Time Image Processing, pp.1-16, 2011.
- [14] J. Canny, "A Computational Approach to Edge Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.PAMI-8, No.6, pp.679-698, 1986.
- [15] <http://djvu.org>
- [16] <http://phenix.int-evry.fr/jct/>

저자 소개

박 상 효(Sang-Hyo Park)

준회원

- 2012년 2월 : 한밭대학교 멀티미디어공학과(공학사)
- 2013년 2월 ~ 현재 : 한밭대학교 멀티미디어공학과 석사 과정

<관심분야> : 영상처리, 영상압축

이 시 웅(Si-Woong Lee)

정회원

- 1997년 8월 : KAIST 전기및전자공학과 (공학박사)
- 1995년 ~ 2000년 : 삼성전자 선임연구원
- 2004년 4월 ~ 현재 : 한밭대학교 정보통신공학과 교수

<관심분야> : 컴퓨터비전, 영상처리, 영상압축