

디지털 비디오에서 문자 영역 이진화를 위한 색상 극화 기법

정 종 면*

The Color Polarity Method for Binarization of Text Region in Digital Video

Jongmyeon Jeong*

요 약

색상 극화란 주어진 텍스트 영역에서 글자 색이 무엇인지를 결정하는 과정으로서 텍스트 추출을 위해서 선행되어야 하는 작업이다. 본 논문에서는 텍스트 영역이 주어졌을 때 글자 영역을 추출하기 위한 색상 극화 기법을 제안한다. 제안된 방법은 글자 영역과 배경 영역에 대한 관찰을 바탕으로 두 영역 사이의 면적 비율과 표준편차 비율의 관계를 색상 극화에 이용한다. 이를 위하여 그레이 스케일로 주어진 텍스트 영역을 Otsu의 방법으로 이진화하고 이진화된 두 영역을 각각 4-CC 레이블링한다. 레이블링된 두 그룹의 영역에 대해 각각 면적과 영역 중심으로부터의 거리에 대한 표준편차를 계산한 다음 두 그룹에서 면적이 가장 넓은 영역을 갖는 영역 사이의 면적 비와 표준편차가 가장 작은 영역들 사이의 표준편차 비를 이용하여 색상 극화를 수행한다. 다양한 폰트와 크기를 갖는 텍스트 영역에 대한 실험을 통해 제안된 방법이 강건하게 색상 극화를 수행함을 확인하였다.

Abstract

Color polarity classification is a process to determine whether the color of text is bright or dark and it is prerequisite task for text extraction. In this paper, we propose a color polarity method to extract text region. Based on the observation for the text and background regions, the proposed method uses the ratios of sizes and standard deviations of bright and dark regions. At first, we employ Otsu's method for binarization for gray scale input region. The two largest segments among the bright and the dark regions are selected and the ratio of their sizes is defined as the first measure for color polarity classification. Again, we select the segments that have the smallest standard deviation of the distance from the center among two groups of regions and evaluate the ratio of their standard deviation as the second measure. We use these two ratio features to determine the text color polarity.

• 제1저자 : 정종면

• 투고일 : 2009. 07. 24, 심사일 : 2009. 08. 12, 게재확정일 : 2009. 08. 31.

* 국립목포해양대학교 해양전자통신공학부 교수

※ 이 논문은 2006년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. KRF-2006-331-D00517).

The proposed method robustly classify color polarity of the text, which has shown by experimental result for the various font and size.

▶ Keyword : 자막 추출(text extraction), 자막 이진화(text binarization), 색상 극화(color polarity)

1. 서 론

컴퓨터 기술의 발전으로 비디오는 방송, 인터넷, 무선 통신망을 통틀어 가장 중요한 미디어 타입이 되었다. 쉽고 빠르게 원하는 비디오 콘텐츠를 찾아내고 사용하기 위한 사용자의 요구가 증대됨에 따라 내용기반 비디오 색인 및 검색 시스템에 대한 연구가 활발히 이루어지고 있는데, 컬러, 모양, 질감, 움직임 정보를 이용하는 것이 일반적이다[1-4].

한편 비디오에 나타나는 텍스트 정보는 비디오 데이터들의 시청각 정보를 보강하고 부가 정보를 사용자에게 제공하며 비디오 데이터의 내용을 예측할 수 있는 단서를 제공할 수 있다. 따라서 내용 기반 비디오 색인 및 검색 시스템을 구축할 때 텍스트 데이터는 중요한 특징으로 사용될 수 있으며, 전통적인 컬러, 모양, 질감, 움직임 정보 등을 이용하는 것에 비해 그 비용과 정확성이 더 우수한 것으로 알려져 있다. 따라서 비디오의 텍스트 정보를 자동으로 추출하기 위한 연구가 활발히 진행되고 있다[5-10].

주어진 비디오에서 텍스트 정보를 추출하는 과정은 텍스트 탐지(text detection), 텍스트 위치 찾기(text localization), 그리고 텍스트 추출(text extraction) 단계로 이루어진다. 텍스트 탐지 단계는 주어진 프레임에 글자가 존재하는지 여부를 개략적으로 판단하는 단계이며 텍스트 위치 찾기는 글자가 존재하는 프레임에서 글자가 정확히 어느 곳에 위치하는지 찾아내는 단계이다. 텍스트 추출 단계는 주어진 텍스트 위치에서 정확하게 글자를 배경으로부터 분리해 내고 그 결과는 OCR에 전달된다[9].

상인지 어두운 색상인지를 판단해야 한다. 그림 1은 텍스트 영역을 이진화한 영상으로써 색상 극화 기법의 필요성을 보이고 있는데, 그림 1(a)는 검정색 영역이 글자영역이고 그림 1(b)는 밝은색 영역이 글자영역이다. 임의의 텍스트 영역에서 자막의 색상이 밝은 색인지 여부는 사전에 알 수 없기 때문에 텍스트 추출 단계에서 자동으로 색상을 극화시킨 다음 두 색상 중 하나를 글자 영역으로 출력해야 한다.

Antani 등은 이진화된 텍스트 영역을 연결 요소(connected-component)로 분류한 다음 글자와 유사한 특징을 갖는 요소에 높은 점수를 부여함으로써 색상을 극화하였다. 이 방법은 글자와 배경 영역이 명확히 구별될 때는 잘 동작하지만 복잡한 배경을 갖는 경우에는 배경이 글자 영역이 요소들과 결합될 가능성이 높아지기 때문에 글자 영역과 배경 영역을 구분짓는 점수에 대한 임계치에 매우 민감하고, 따라서 복잡한 배경을 갖는 경우에는 적용하기 힘들다[11]. Lienhart 등은 칼라 히스토그램에 기반한 색상 극화 기법을 제안하였는데, 텍스트 영역의 중앙과 위아래 영역의 색상 분포를 검사하였다. 이 방법은 임계치를 설정할 필요없이 없다는 장점이 있으나(점수배경과 글자 영역이 명확히 구별될 때 잘 동작한다. 그러나 글자 영역과 배경 영역이 비슷한 색상일 경우에는 동작하지 못한다[12]. Song 등은 텍스트 색상에 대한 통계적 분석을 통해 색상을 극화하는 기법을 제안하였다. 이 방법은 이진화된 텍스트 영역에서 칼라 에지와 배경 에지의 관계를 분석하여 글자 영역과 배경 영역의 에지경울만들[11] 두 개의 비율을 통계적으로 분석한 다음 이 경험적(heuristic) 임계치로 사용하여 글자와 배경의 색상을 극화하였다. 이 방법은 글자 영역 하위의 색상 분포에 무관화 기법을 제안하였는데, 적용할 수 있지만, 다양한 폰트에 대한 통계적 분석이 어렵고 폰트 하위에 특수효과가 있거나 복잡한 배경이 있고할 경우에는 오동작하기 쉽다는 단점이 있다[13-14]. Jung 등은 한 텍스트 영역에 서로 다른 색상이 하고, 따경우의 색상 극화 제위하여 이중 이진화(du글자 binarization) 기법을 제안하였다. 이 방법은 그레이 스케일의 영상을 Otsu 방법을 이용하여 이진화한 다음, 어두운 영역과 밝은 영역 각각에 대해서 미리상 의된 글자 크기보다작은 영역을 제거한 다음 다시 어두운 영역과 밝은 영역을 결합하여 최종 중앙 제출출하였는데, 글자의 크기 제미리상 의하기가 어렵고 복잡한 배경이 있을 경우 동작하기 힘들다[15]. Jung 등은 스트록 필터(stroke f객자ter) 제어용한

지역사무소장

(a) 어두운 글자의 예

생각할 건

(b) 밝은 글자의 예

그림 1. 색상 극화의 필요성

Fig 1. Necessity of color polarity

색상 극화(color polarity) 기법은 텍스트 추출 단계에서 텍스트 영역을 이진화(binanzation) 시켰을 때 글자와 배경 영역을 구분하기 위한 기법이다. 이진화된 텍스트 영역은 글자와 배경 영역으로 나누어지는데, 이때 글자 영역의 밝은 색

텍스트 추출 방법을 제안하였다. 스트로크 필터 매우 민감어둡거나 밝은 경우 제모두 고려한 필터로서, 스트로크와 유사한 구조 제깃는 밝기 분포에서 높은 응답이 나오도록 설계되었다. 우 민감밝을 경우 밝은 스트로크에 대한 필터 응답이 어두운 스트로크에 대한 응답보다 크 기나타난다는 점을 이용하여 두 응답에 대한 비율을 계산하고 또한 이진화된 텍스트 영역에서 어두운 에지와 밝은 에지의 비율을 계산한 다음 스트로크 필터의 비율과 에지의 비율을 SVM(Support Vector 읍한 hine) 분류기 제이용하여 우 민감어두운지 밝은지를 판단하였다. 이 방법은 어두운 배경에 계되었다.자가 있는 경우, 또는 밝은 배경에 밝은 글자가 있는 경우에도 잘 동작하지만, 배경이 복잡하거나, 글자와 유사한 구조가 배경 영역에 존재할 경우에는 오동작할 수 있다[16].

본 논문에서는 복잡한 배경을 갖는 환경에서도 강건하게 동작할 수 있는 색상 극화 방법을 제안한다. 제안된 방법은 비디오 텍스트 영역에 대한 다수의 실험과 관찰을 바탕으로 하고 있다. 본 논문의 이후 구성은 다음과 같다. 먼저 II 장에서 제안하는 방법의 개요와 이진화 과정을 설명한 다음 면적 비율을 이용한 색상 극화 척도와 표준편차 비율을 이용한 색상 극화 척도를 설명하고 이를 이용한 색상 극화 기법을 기술한다. III 장에서 실험 결과를 보인 후 IV장에서 결론을 맺는다.

II. 제안하는 방법

1. 개요

본 논문에서 제안하는 방법은 자막 위치 찾기(text localization) 단계에서 얻어진 그레이 스케일을 갖는 텍스트 영역을 입력으로 받아 글자 영역을 밝게 배경 영역을 어둡게 하는 이진화된 영상을 출력으로 한다. 글자와 배경 사이의 구조적 특징을 고려하여 색상을 극화시키는데, 이를 위하여 먼저 주어진 그레이 스케일의 영상을 Otsu 의 방법으로 이진화하여 밝은 영역과 어두운 영역으로 분할한다. 연결 요소(connected-component) 분석을 통해 밝은 영역은 밝은 연결 요소를 갖는 영역들로, 어두운 영역은 어두운 연결 요소들을 갖는 영역들로 분할한 다음, 분할된 밝은 영역과 어두운 영역 중 그 크기가 가장 큰 영역의 면적들 사이의 비율과, 두 영역 중 영역의 분포가 가장 압축되어 있는 영역들의 압축된 정도 사이의 비율, 그리고 두 영역 사이의 포함관계를 검사하여 색상을 극화 시킨다. 그림 2는 제안된 방법을 개략적으로 보여주는 블록도이다.

2. 이진화

제안된 방법은 텍스트 위치 찾기(text localization) 단계에서 찾아진 텍스트 영역을 그레이 스케일로 입력 받는다. 주어진 영상을 이진화 하기 위하여 본 논문에서는 Otsu 등의 방법을 사용하는데, Otsu 등의 방법은 그레이 스케일 영상의 색상 분포에 두 개의 큰 산(peak)와 한 개의 계곡(vally)가 존재할 때 잘 동작하는 이진화 알고리즘으로 알려져 있다[17]. 디지털 비디오에서 텍스트 영역은 글자 부분과 배경 영역이 가시성을 향상시키기 위하여 큰 대비(contrast)를 갖기 때문에 Otsu 등의 이진화 알고리즘은 텍스트 영역에 대한 이진화를 위해 적합하다고 할 수 있다. 한편 본 논문에서는 그레이 스케일로 주어진 입력 영상을 상하좌우로 획 두께(stroke width) 정도의 크기만큼 확장한 다음 이진화를 수행하였으며 이진화된 영상에서 글자 영역은 밝은 영역일 수도 어두운 영역이 될 수도 있다. 이진화된 영상에서 밝은 영역은 밝은 영역대로 4-CC 레이블링을 통해 분할하고 어두운 영역은 어두운 영역대로 4-CC 레이블링을 수행하여 분할한다.

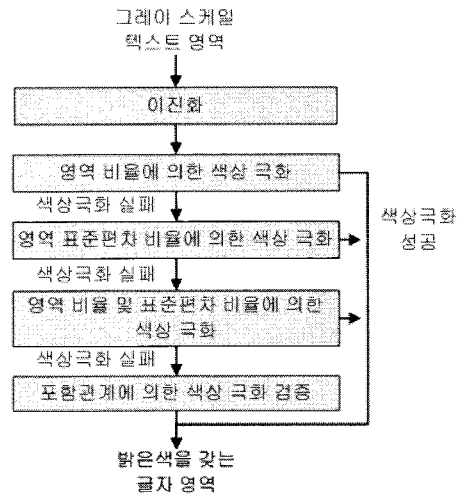


그림 2. 제안된 방법의 블록도
Fig. 2. Block diagram of the proposed method

3. 면적 비율을 이용한 척도

텍스트 영역은 글자와 배경으로 이루어진다. 글자는 획(stroke)들의 집합으로 구성되고 배경은 획의 바깥 영역을 둘러싸는 영역과 획으로 둘러싸인 좁은 영역들로 이루어지는데, 그림 3에서 보이는 바와 같이 배경 중 가장 큰 면적을 갖

는 영역은 모든 픽들을 둘러싸는 영역이 될 가능성이 높다.

그림 3(a)는 원영상의 예를 보이고 있으며 그림 3(b)는 원 영상을 밝은 영역과 어두운 영역들 중 가장 큰 면적을 갖는 영역을 표시하고 있는데, 검정색으로 표현된 영역이 어두운 영역 중 가장 넓은 영역을 보이고 있고, 흰색으로 보이는 영역이 밝은 색 영역 중 가장 넓은 영역을 보이고 있다. 연두색으로 표현된 부분은 나머지 영역이다.

그림 3은 텍스트 영역에서 글자와 배경 영역의 면적에 색상 극화를 위한 중요한 정보가 될 수 있음을 보이고 있다. 본 논문에서는 밝은 영역과 어두운 영역에서 가장 넓은 면적을 갖는 두 영역 사이의 면적을 계산하고 이 면적들 사이의 면적 비율을 계산하여 색상 극화에 사용한다. 이를 위하여 텍스트 영역에서 밝은 영역과 어두운 영역을 각각 4-CC 레이블링한 후 두 영역에서 가장 큰 면적을 갖는 영역을 선택한다. 밝은 영역에서 가장 큰 영역의 면적이 BR 이라 하고 어두운 영역에서 가장 큰 영역의 면적이 DR 이라고 하면 정규화된 면적의 차이의 비율 R_{ratio} 는 식 1과 같다.

$$R_{ratio} = (BR - DR) / \text{MAX}(BR, DR) \dots\dots\dots (\text{식 } 1)$$

위 식 1에서 어두운 배경에 밝은 글자가 존재할 경우 R_{ratio} 는 음수로 나타나며, 그 반대의 경우에 R_{ratio} 는 양수로 나타나는데, 그 절대값이 1에 근접할수록 신뢰할 수 있는 색상 극화 결과를 얻을 수 있다.



그림 3. 최대 면적을 갖는 글자 영역과 배경 영역
Fig. 3. Text and background regions with the largest size

4. 표준편차 비율을 이용한 척도

면적 비율을 이용한 색상 극화는 텍스트 영역을 표현하는 비운딩박스가 완벽하게 글자를 포함하고 있을 경우에 좋은 성능을 보일 수 있다. 그러나 자막 위치찾기 단계에서 텍스트 영역이 글자의 일부에 겹치도록 얻어지거나 또는 복잡한 배경의 영향으로 배경 자체가 밝고 어두운 색을 모두 포함할 경우에는 배경 영역은 여러 개로 나누어지고 그 결과 면적 비율을 이용한 색상 극화는 신뢰할 수 있는 결과를 얻어내기 힘들다. 텍스트 영역을 글자와 배경 영역으로 각각 분류했을 때 색상을 극화시키기 위한 또 하나의 척도가 될 수 있는 것이 영역의 흠어짐 정도이다. 글자는 한 개 이상의 스트로크로 이루어지는데 스트로크는 배경 영역과 비교했을 때 공간적으로 더 퍼져있는 경우가 많다.

본 논문에서는 글자 스트로크가 내부에 일정한 배경 영역 즉, 구멍(hole)을 포함하는 경우를 주목하였다. 그림 3(a)의 첫 번째, 세 번째, 네 번째, 여섯 번째, 일곱 번째의 텍스트 영역은 스트로크가 자신의 내부에 배경 영역을 포함하고 있다. 즉, 첫 번째는 “o”이, 세 번째는 “a”와 “g”가 네 번째는 “b”와 “e”가 여섯 번째는 “o”과 “r”이 일곱 번째는 “교”이 자신의 내부에 배경 영역을 포함하고 있는데, 그림에서 보이는 것처럼 스트로크 내부의 배경 영역은 공간적으로 매우 뭉쳐있다. 본 논문에서는 밝은 색과 어두운 색의 각 영역의 중심 좌표를 구하고, 그 중심좌표로부터 영역 내 픽셀의 거리들의 표준편차를 식 2 - 식 5와 같이 구한다.

$$(x_m, y_m) = \left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n y_i \right) \dots\dots\dots (\text{식 } 2)$$

$$\text{dist}(i) = \sqrt{(x_i - x_m)^2 + (y_i - y_m)^2} \dots\dots\dots (\text{식 } 3)$$

$$m_{dist} = \left(\frac{1}{n} \sum_{i=1}^n \text{dist}(i) \right) \dots\dots\dots (\text{식 } 4)$$

$$S_{dist} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n ((m_{dist} - \text{dist}(i))^2)} \dots\dots\dots (\text{식 } 5)$$

여기에서 (x_i, y_i) 는 주어진 영역에 존재하는 한 좌표를 의미하고, (x_m, y_m) 는 주어진 영역의 중심좌표를, n은 영역 내 픽셀의 개수를 의미한다. 또한 $\text{dist}(i)$ 는 영역의 중심좌표와 i 번째 점의 유클리디언 거리를 의미하며 m_{dist} 는 $\text{dist}(i)$ 의 평

균을, S_{abs} 는 영역의 중심좌표로부터 영역 내 픽셀 거리의 표준편차를 각각 의미한다.

이진화된 텍스트 영역에서 밝은 영역 중 가장 표준편차가 작은 영역을 선택하고, 동일한 방법으로 어두운 영역 중 가장 표준편차가 작은 영역을 선택한다. 두 영역의 표준편차 중 어느 한쪽의 표준편차가 작고, 나머지 한쪽의 표준편차가 매우 큰 값을 갖는다면 표준편차가 작은 영역은 스트로크 내부에 있는 배경 영역일 가능성이 높다. 한편 배경 영역이 스트로크 보다 더 넓게 퍼져있는 경우가 있으나, 이 경우는 그림 3에서 보이는 바와 같이 배경이 글자를 둘러싼 경우를 의미하기 때문에 배경 영역의 면적이 스트로크 영역에 비해 훨씬 크며, 따라서 이럴 경우는 면적 비율 계산에 의해 신뢰할 수 있는 색상 극화를 수행할 수 있다.

식 5에 의해 계산된 각 영역들의 표준편차들 중 밝은 영역 중에서 가장 작은 표준편차 값을 갖는 영역의 표준편차를 BS 라 하고 어두운 영역 중 가장 작은 표준편차 값을 갖는 영역의 표준편차를 DS 라고 한다면 표준편차 비율 S_{ratio} 는 다음과 같이 계산된다.

$$S_{\text{ratio}} = (DS - BS) / \text{MAX}(BS, DS) \dots\dots\dots (\text{식 } 6)$$

위 식에서 어두운 배경에 밝은 글자가 존재할 경우 S_{ratio} 는 음수로 나타나며, 그 반대의 경우에 S_{ratio} 는 양수로 나타내는데, 그 절대값이 1에 근접할 수록 신뢰할 수 있는 색상 극화 결과를 얻을 수 있으며 0에 근접할수록 그 결과를 신뢰하기 힘들다.

5. 색상 극화 및 검증

표준편차 비율을 이용한 색상 극화는 스트로크 내부에 배경을 포함하고 있을 경우 대체적으로 좋은 성능을 보일 수 있다. 그러나 한글이나 한문의 경우 폰트에 따라, 영어의 경우 i, j와 같이 점(dot)이 획의 일부인 경우 또는 쉼표(.)나 마침표(.)와 같은 특수 문자가 텍스트 영역에 포함되는 경우는 글자 영역에서도 표준편차가 매우 작게 나타난다. 또한 배경이 극단적으로 복잡하여 배경에 글자색과 같은 색상과 그 반대의 색상이 번갈아 나올 경우에는 면적 비율과 표준편차 비율의 절대값이 크지 않고 그 결과 신뢰하기 어려운 결과가 나올 수 있으며 이럴 경우 S_{ratio} 의 절대값은 그리 크지 않다. 따라서 면적 비율과 표준편차 비율 척도가 모두 임계치보다 작을 경우 본 논문에서는 두 비율에 대한 가중치가 부여된 합을 구하여 색상 극화를 수행한다.

$$Polarity_{\text{ratio}} = w_1 R_{\text{ratio}} + w_2 S_{\text{ratio}} \dots\dots\dots (\text{식 } 7)$$

여기에서 w_1, w_2 는 가중치이며 $w_1 + w_2 = 1$ 이고 $Polarity_{\text{ratio}}$ 는 (-1, 1)의 범위를 갖는다. 다른 척도들과 마찬가지로 $Polarity_{\text{ratio}}$ 의 절대값이 1에 근접할 수록 신뢰할 수 있는 결과를 얻을 수 있으며 0에 근접할 수록 색상의 선택이 모호한 경우이다. 따라서 $Polarity_{\text{ratio}}$ 의 절대값이 임계치보다 크면 $Polarity_{\text{ratio}}$ 를 이용하여 색상 극화를 수행하고, 면적 비율과 색상 비율로도 신뢰할 수 있는 결과를 보이기 힘들 경우 포함 관계를 이용한 검증을 통해 최종적인 결과를 도출한다.

본 논문에서 제안하는 색상 극화 기법을 요약하면 다음과 같다. 본 논문에서는 다수의 실험을 통한 관찰을 바탕으로 배경과 글자 영역 사이의 관계를 정량적인 수치로 표현하고 그 수치에 따라 색상 극화를 수행한다. 이를 위하여 먼저 주어진 텍스트 영역을 이진화 한 후 밝은 영역과 어두운 영역에 대한 면적 비율 R_{ratio} 와 표준편차 비율 S_{ratio} 를 계산한다. 실험을 통해 관찰한 결과 대부분의 경우 면적 비율이 표준편차 비율 보다 더 신뢰할 수 있는 결과를 보이기 때문에 면적 비율의 절대값이 임계치보다 클 경우 면적 비율 척도만으로 색상 극화를 수행한다. 면적 비율의 절대값이 임계치보다 작고 표준편차 비율의 절대값이 임계치보다 클 경우에는 표준편차 비율만 이용하여 색상 극화를 수행한다. 면적 비율과 표준편차 비율의 절대값이 모두 임계치보다 작을 경우에는 면적 비율과 표준편차 비율에 가중치가 부여된 합을 이용하여 색상 극화에 사용하여, 포함관계를 이용하여 이를 검증한다.

III. 실험 결과

본 논문에서 제안된 알고리즘의 타당성을 보이기 위해 제안된 알고리즘을 Visual C/C++ 2008 환경에서 구현하였다. 실험에 사용한 영상들은 국내외에서 방송된 비디오에서 얻은 것들이며 쇼, 뉴스, 스포츠, 다큐멘터리, 영화 등을 망라한다.

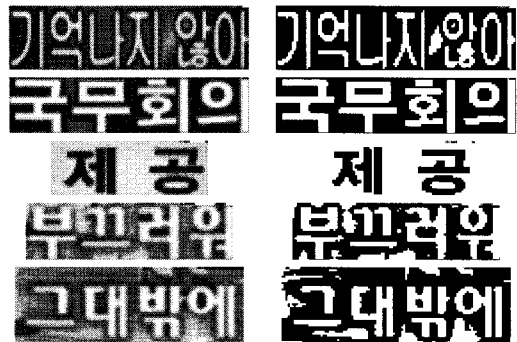




그림 4 입력 영상
Fig 4 Input Images

입력 비디오로부터 텍스트 영역을 3가지 형태로 추출하여 실험에 사용하였는데, 텍스트 추출 단계에서의 다양한 오류를 고려하기 위하여 정확한 텍스트 영역이 추출된 경우와 함께 텍스트 영역보다 더 많은 배경을 포함하는 경우, 텍스트 영역이 좁게 추출되어 텍스트 영역의 일부가 없는 경우로 나누어 실험하였다.

그림 4는 실험에 사용한 영상들 중 텍스트에 테두리 효과나 그림자 효과가 있거나 배경이 복잡한 열악한 예들을 보이고 있다. 그림 4(a)는 입력 텍스트 영상을 보이고 있는데, 그림 4(a)의 첫 번째 영상은 텍스트에 테두리 효과가, 두 번째 영상은 텍스트에 그림자 효과가, 세 번째 영상은 밝은 색 배경에 어두운 글자가 존재하는 예를 각각 보이고 있다. 또한 네 번째, 다섯 번째, 여섯 번째 영상들은 배경이 극단적으로 복잡한 경우의 예를 보이고 있으며 일곱 번째와 여덟 번째 영상은 영어와 일본어에 대한 예를 보이고 있다. 그림 4(b)는 그림 4(a)의 입력 영상을 그레이 스케일 영상으로 변환 한 후 Otsu의 방법으로 각각 이진화한 결과이다. 그림 5는 그림 4(b)의 이진화된 텍스트 영상에 대해 색상 극화를 수행한 결과를 단계별로 보이고 있는데, 그림 5(a)는 이진화된 텍스트 영상에서 밝은 영역과 어두운 영역 중 가장 넓은 영역들을 보이고 있다. 여기에서 밝은 영역 중 가장 넓은 영역이 흰색으로, 어두운 영역 중 가장 넓은 영역이 검정색으로 표현되었으며 녹색 영역은 나머지 영역을 각각 의미한다.

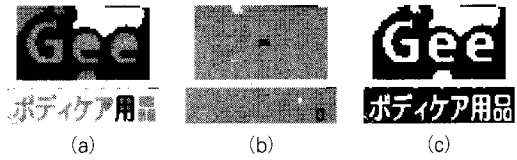
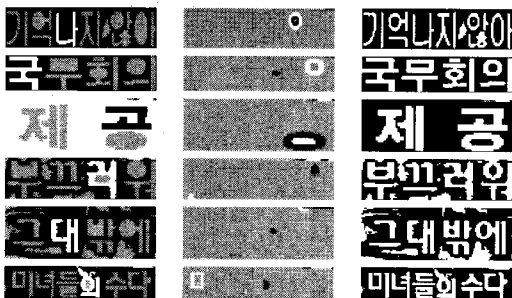


그림 5 실험 결과
Fig. 5 Experimental results

그림 5(b)는 밝은 색 영역과 어두운 색 영역 중 표준편차가 가장 작은 영역을 보이고 있는데, 밝은 색 영역 중 가장 표준편차가 작은 영역이 흰색, 어두운 영역 중 가장 표준편차가 작은 영역이 검정색, 그리고 회색 영역은 그 나머지 영역을 의미한다. 마지막으로 그림 5(c)는 최종 자막 추출 결과를 보이고 있는데, 흰색 영역이 글자 영역, 그리고 검정색 영역이 배경 영역을 의미한다. 본 논문에서 제안한 방법으로 다양한 폰트와 크기를 갖는 501개의 텍스트 영역에 대해 실험을 수행하였고 그 결과 정확하게 텍스트 영역이 입력된 경우와 실제 텍스트 영역보다 더 큰 텍스트 영역이 주어졌을 경우에는 98.8%의 정확도로 색상극화를 수행함을 확인할 수 있었으며 텍스트 영역이 실제 텍스트보다 작게 주어졌을 경우 97.6%의 정확도로 색상 극화를 수행함을 확인하였다. 한편 기존의 색상 극화 알고리즘 중 대표적인 알고리즘이라 할 수 있는 Song 등의 방법[13]에서는 200개의 실험데이터에 대하여 98.5%의 정확도로 색상 극화가 수행되었음을 보이고 있는데, 실험에 사용한 데이터의 출처가 공개되지 않아 본 논문에서 제안된 방법과의 의미 있는 비교는 어렵다. 특히 그들의 방법은 텍스트의 에지에 기반하고 있기 때문에 텍스트이 정확하게 주어졌다는 가정이 필요하며, 그 결과 본 논문에서 실험한 것과 같이 텍스트 영역이 실제 텍스트 영역보다 작게 주어졌을 경우에는 신뢰할 수 있는 결과를 얻기 힘들다.

표 1. 실험 결과
Table 1. Experimental result

	테스트 영상의 수	올바른 색상 극화의 수	정확도
큰 텍스트 영역	167	165	0.988
정확한 텍스트 영역	167	165	0.988
작은 텍스트 영역	167	163	0.976
계	501	493	0.984

IV. 결 론

본 논문에서는 텍스트 영역이 주어졌을 때 글자를 추출하기 위한 색상 극화 알고리즘을 제안하였다. 제안된 방법은 글자 영역과 배경 영역에 대한 관찰을 바탕으로 두 영역 사이의 면적 비율과 표준편차 비율의 관계를 색상 극화에 이용한다. 이를 위하여 먼저 주어진 텍스트 영역을 Otsu의 방법으로 이진화하고 이진화된 두 영역을 각각 4-CC 레이블링한다. 레이블링된 두 그룹의 영역에 대해 각각 면적과 영역 중심으로부터의 거리에 대한 표준편차를 계산한 다음 두 그룹에서 면적이 가장 넓은 영역을 갖는 영역 사이의 면적 비와 표준편차가 가장 작은 영역들 사이의 표준편차 비를 이용하여 색상 극화를 수행한다. 제안된 알고리즘을 다양한 폰트와 크기의 글자에 대해 실험한 결과 정확하게 텍스트 영역이 주어진 경우와 텍스트 영역이 크게 주어진 경우에는 98.8%의 정확도로, 텍스트 영역이 작게 주어진 경우에는 97.6%의 정확도로 색상 극화를 수행함을 확인하였다. 향후 밝은 배경에 밝은 글자가 존재하는 경우, 또는 어두운 배경에 어두운 글자가 존재하는 경우에 대한 연구가 필요하다.

참고문헌

- [1] A. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. on Pattern Analysis and Machine Intelligent*, vol. 22, no. 12, pp. 1349-1380, Dec. 2000.
- [2] 박기태, 문영식, "색상의 공간적인 상관관계와 국부적인 푸리에 변환에 기반한 질감 특성을 이용한 영상 검색," *대한전자공학회논문지*, 제 44권 SP편 제 1호, 10-16 쪽, 2007년 2월.
- [3] 정명범, 김재경, 고일주, 장대식, "특수 영상에서 비디오 요약을 위한 장면 전환 검출 알고리즘," *한국컴퓨터정보학회논문지*, 제 14권, 제 3호, 65-74쪽, 2009년 3월.
- [4] 신성윤, 표성배, 이양원, "대용량 비디오 데이터베이스 구축을 위한 비디오 개요 추출," *한국컴퓨터정보학회논문지*, 제 14권, 제 1호, 255-265쪽, 2006년 3월.
- [5] A. Jain and B. Yu, "Automatic text location in images and video frames," *Pattern Recognition*, vol. 31, no. 12, pp. 2055-2076, 1998.

- [6] Y. Hasan and L. Karam, "Morphological Text Extraction from Images," *IEEE Trans. on Image Processing*, vol. 9, no. 11, Nov. 2000.
- [7] C. W. Lee, K. Jung, H. J. Kim, "Automatic text detection and removal in video sequences," *Pattern Recognition Letters*, vol. 24, pp. 2607-2623, Nov. 2003.
- [8] 정종면, 차지훈, 김규현, "디지털 비디오를 위한 획기반 자막 추출 알고리즘," *퍼지 및 지능시스템학회 논문지*, 제 17권, 제 3호, 297-303쪽, 2007년 6월.
- [9] K. Jung, K. Kim, A. Jain, "Text Information extraction in images and video: a survey," *Pattern Recognition*, vol. 37, pp. 977-997, May 2004.
- [10] V. Wu, R. Manmatha, and E. Riseman, "Textfinder: An automatic system to detect and recognize text in images," *IEEE Trans. on Pattern Analysis and Machine Intelligent*, vol. 21, no. 11, pp. 1224-1229, Nov. 1999.
- [11] S. Antani, D. Crandall, and R. Kasturi, "Robust extraction of text in video," *15th Int. Conf. on Pattern Recognition*, vol. 1, pp. 831-834, Sep. 2001.
- [12] R. Lienhart and A. Wernicke, "Localizing and Segmenting Text in Images and Videos," *IEEE Trans. on Circuits and Systems for Video technology*, vol. 12, no. 4, pp. 256-268, Apr. 2002.
- [13] J. Song, M. Cai and M. R. Lyu, "A Robust Statistic Method for Classifying Color Polarity of Video Text," *Proc. of IEEE Int. Conf. of Acoust., Speech, Signal Process*, vol. 3, pp. 581-584, 2003.
- [14] M. R. Lyu, J. Song, and M. Cai, "A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction," *IEEE Trans. on CSVT.*, vol. 15n no 2, Feb. 2005.
- [15] C. Jung and J. Kim, "Player Information Extraction for Semantic Annotation in Golf Videos", *IEEE Trans. on Broadcasting*, vol. 55, no. 1, pp.79-83, Mar. 2009.

- [16] C. Jung, Q. Liu, J. Kim, "A New Approach for Text Segmentation Using a Stroke Filter", International Journal of Signal Processing, vol 88, Issue 7, pp. 1907-1916, 2008.
- [17] N. Otsu, "A Threshold Selection method from Gray-Level Histograms", IEEE Trans. on Syst., Man, Cybern., vol. 9, no. 1, pp. 62-66, Mar. 1979.

저자 소개



정 증 면

1992 : 한양대학교 공학사
1994 : 한양대학교 공학석사
2001 : 한양대학교 공학박사
2001 - 2004 : 한국전자통신연구원
 신입연구원
2008 - 2009 :
The Ohio State University
Visiting Scholar
2004 - 현 재 :
국립목포해양대학교
해양전자통신공학부 교수
관심분야 : 영상처리, 머신 비전, 디
 지털방송, MPEG-2, 4,
 7, 21 응용