

# 자연영상 텍스트 이진화를 위한 3단계 색상 군집화 알고리즘

김 지 수<sup>†</sup> · 김 수 형<sup>††</sup>

## 요 약

본 논문에서는 자연 영상에서 추출된 텍스트 영상의 이진화를 위한 3단계 색상 분할 알고리즘을 제안한다. 1단계 색상 분할은 입력 영상 중에서 텍스트색상과 배경색상이 잘 구별되는 단순 영상의 이진화를 수행하고, 2단계 색상 분할은 입력 영상에 고주파 필터를 적용하여 자연조명이나 인공조명에 영향을 받은 영상의 이진화를 수행하며, 3단계 색상 분할은 저주파 필터를 기반으로 텍스트나 배경에 텍스처가 존재하는 영상의 이진화를 수행한다. 제안한 이진화 알고리즘이 그레이 정보를 이용하는 이진화 알고리즘보다 효과적으로 텍스트 영역을 이진화함을 입증하기 위해 상용문서 인식기인 아르미 6.0을 사용하여 이진화된 문자영상에 대한 인식 실험을 실행하였다. 그 결과 제안한 알고리즘의 필드 단위 인식률이 그레이 정보를 이용한 이진화 알고리즘의 인식률보다 35% 이상 우수함을 판측하였다.

**키워드 :** 장면 텍스트 인식, 적응적 이진화, 색상 군집화

## Three-Level Color Clustering Algorithm for Binarizing Scene Text Images

Kim Ji-Soo<sup>†</sup> · Kim Soo-Hyung<sup>††</sup>

## ABSTRACT

In this paper, we propose a three-level color clustering algorithm for the binarization of text regions extracted from natural scene images. The proposed algorithm consists of three phases of color segmentation. First, the ordinary images in which the texts are well separated from the background, are binarized. Then, in the second phase, the input image is passed through a high pass filter to deal with those affected by natural or artificial light. Finally, the image is passed through a low pass filter to deal with the texture in texts and/or background. We have shown that the proposed algorithm is more effective used gray-information binarization algorithm. To evaluate the effectiveness of the proposed algorithm, we use a commercial OCR software ARMI 6.0 to observe the recognition accuracies on the binarized images. The experimental results on word and character recognition show that the proposed approach is more accurate than conventional methods by over 35%.

**Key Words :** Scene Text Recognition, Adaptive Binarization, Color Clustering

### 1. 서 론

일상생활에서 많이 사용하고 있는 핸드폰 카메라, 디지털 카메라 그리고 디지털 캠코더로 취득된 정지영상이나 비디오 프레임에 인위적으로 삽입되거나 자연적으로 포함된 텍스트들은 영상내의 내용을 함축적이고 구체적으로 표현하는 중요한 정보들이다. 이러한 정보들을 실시간에 추출하여 인식 할 수 있다면 로봇 자동 주행 시스템, 시각 장애인 주행 보조 시스템, 고용량 비디오 프레임의 자동 검색 및 색인 시스템, 텍스트 자동 번역 시스템 등과 같은 다양한 분야에 널리 활용될 수 있다.

지금까지의 장면 텍스트 처리 연구는 크게 두 가지로 나누어 볼 수가 있다. 첫 번째 연구는 영상 내에 포함된 텍스

트 영역의 위치를 찾아서 추출해내는 텍스트 후보 영역 추출에 관한 연구이고[1-6, 13-18], 두 번째는 추출된 텍스트 영상을 이진화(binazation) 후 OCR(Optical Character Recognition)을 사용하여 인식하는 연구이다[7-12].

정지영상이나 연속적인 비디오 프레임으로부터 추출된 텍스트 영상을 이진화한 후 OCR을 사용하여 인식하는 관련 연구를 살펴보면 <표 1>과 같다. 영상에 존재하는 텍스트를 인식하기 위해서는 전처리 작업으로 이진화가 선행되어야 하며, 이진화 결과는 OCR 인식 결과에 많은 영향을 주게 된다. <표 1>의 관련 연구들은 추출된 텍스트 영상을 이진화 할 때에 대부분 컬러 영상을 그대로 이용하지 않고 그레이 영상으로 변환하여 이진화한다. 이와 같이 컬러 영상을 그레이 영상으로 변환한 후에 이진화를 수행하게 되면 컬러 정보를 손실하게 되어 이진화 결과에 나쁜 영향을 주게 되는 단점이 존재한다.

본 논문에서는 기존 연구의 단점을 해결하기 위해서 영상

※ 이 논문은 2005년도 전남대학교 학술연구비 지원에 의하여 연구되었음.

† 준 회 원 : 전남대학교 전산학과 박사과정

†† 정 회 원 : 전남대학교 전자컴퓨터정보통신공학부 부교수

논문접수 : 2005년 9월 16일, 심사완료 : 2005년 11월 8일

<표 1> 자연영상 내 텍스트 이진화 관련연구

참고문헌	입력영상	이진화 방법	문자인식기
Hori 등[7]	그레이 비디오 프레임	입력 영상의 히스토그램 및 임계값을 이용	상용 문서 인식기
Ohya 등[8]	그레이 정지영상	8*8 크기의 작은 블록의 로컬 임계값을 이용	상용 문서 인식기
Wang 등[9]	컬러 정지영상	위상학적 특징, 유클리디안 거리 및 Modified Coarse-Fine Fuzzy C-Means 알고리즘을 이용	상용 문서 인식기
Wolf 등[10]	그레이 비디오 프레임	Niblack's 알고리즘을 이용.	Finereader 5.0
Wu 등[11]	컬러영상과 그레이영상	입력 영상의 히스토그램 및 임계값을 이용	Caere's Omnipage Pro 8.0
Zhang 등[12]	그레이 정지영상	명암 정규화, 가버(gabor) 필터, 선형 판별 분석기를 이용	자체 개발 인식기

자체가 갖고 있는 정보의 손실이 없는 3단계 색상 분할 방법을 사용한 이진화 방법을 제안한다.

2. 제안 방법

본 논문에서는 기존 연구의 단점인 영상 변환 작업에 의해서 나타나는 정보의 훼손을 막기 위해서 컬러 영상이 갖고 있는 정보를 최대한 훼손하지 않는 3단계 색상 분할 방법을 사용한 이진화 방법을 제안한다. 제안한 시스템은 (그림 1)에서와 같이 3단계 과정으로 이진화를 수행함으로써 자연 영상이 갖는 다양한 색상, 조명, 텍스처에 영향을 받지 않고 이진화를 수행할 수 있다.

본 논문의 실험에 사용한 입력 영상들은 우리 일상생활에서 흔히 접하게 되는 학교, 병원, 지하철역, 도로 간판, 상가 간판 및 현수막 등의 실내외에서 취득한 120개의 자연 영상(KAIST 및 숙명여대에서 수집)과 2003 ICDAR Contest Image (Training용 116개 및 Test용 251개 등 총 487개 영상)등 총 607개 자연영상을 대상으로 [18] 시스템을 적용하여 추출한 텍스트 영상들이다. 이러한 영상들은 크게 3가지로 분류할 수 있다.

첫째, 텍스트 컬러와 배경 컬러가 잘 구별되는 단순 영상  
 둘째, 자연조명이나 인공조명에 의해서 영향을 받은 영상  
 셋째, 텍스트나 배경에 텍스처가 존재하는 영상

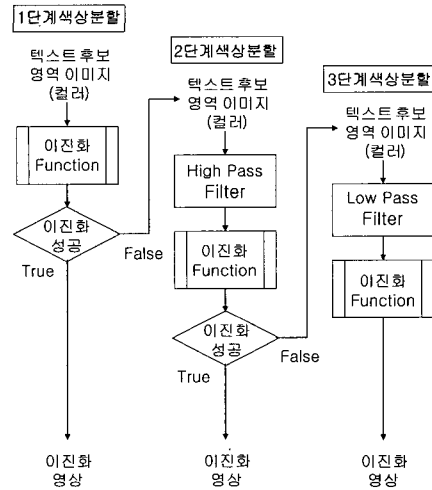
(그림 2)는 다양한 입력 영상을 보여주고 있다. <표 2>와 <표 3>은 자연 영상 및 2003 ICDAR 영상 그리고 이 영상들로부터 추출한 입력 영상의 크기 및 영상 내에 존재하는 텍스트의 크기를 정리하였다.

<표 2> 자연 영상 및 2003 ICDAR 영상 분석 결과

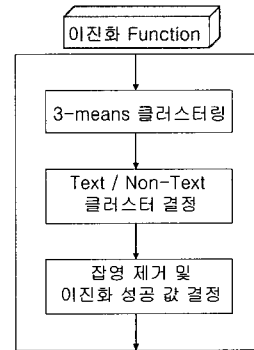
영상 가로 크기(화소)	320 - 1600(화소)
영상 세로 크기(화소)	320 - 1600(화소)
컬러 값	24bit Color
이미지 해상도(dpi)	72 - 96(dpi)

<표 3> 입력 영상의 크기 및 텍스트 크기

입력 영상의 크기 및 해상도			영상내의 텍스트 크기	
세로높이	가로넓이	해상도	텍스트 높이	텍스트 넓이
11-180 화소	30-280 화소	72dpi	7-170 화소	7-270 화소



(ㄱ) 알고리즘 개요



(ㄴ) 이진화 함수

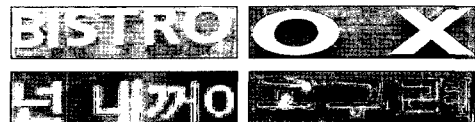
(그림 1) 제안한 이진화 알고리즘



(ㄱ) type 1 영상



(ㄴ) type 2 영상



(ㄷ) type 3 영상

(그림 2) 입력 영상의 예

2.1 1단계 색상 분할

1단계 색상 분할은 입력 영상 중에서 텍스트색상과 배경 색상이 잘 구별되는 단순 영상의 이진화를 목표로 한다. 1

단계에서는 전처리 없이 이진화를 수행함으로써 다음과 같은 장점이 있다. 첫째, 영상내의 정보 훼손이 없기 때문에 최적의 이진화 결과를 얻을 수 있다. 둘째, 빠른 이진화가 가능하며, 초기에 적은 노력으로 많은 효과를 얻을 수 있다.

2.1.1 이진화 Function

가. 3-means 클러스터링

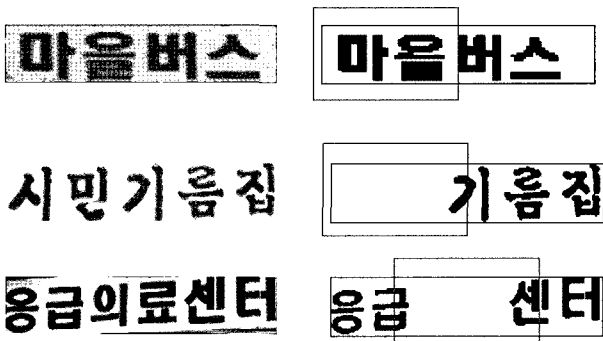
클러스터링(clustering)이란 서로 간에 높은 유사도를 갖는 객체들을 같은 클러스터로 그룹화하는 과정으로, 크게 분할 방법(partitioning method), 계층적 방법(hierarchical method), 밀도기반 방법(density-based method), 그리드기반 방법(grid-based method), 그리고 모델기반 방법(model-based method)의 다섯 가지 종류로 나눌 수 있다[22]. 본 논문에서는 일반적으로 자주 사용되는 분할 방법의 일종인 k-means 방법을 채택하여 문자와 배경을 구분하는데 적용하고자 한다. k-means 알고리즘은 객체를 k개의 그룹으로 분할하는 방법으로, 이때 클러스터에 속하는 객체들의 평균값(중심점)을 사용한다.

일반적으로 텍스트 후보 영상내의 클러스터는 텍스트 클러스터와 배경 클러스터의 2종류로 나눌 수 있다. 이렇게 두 개의 클러스터를 가정했을 때 크기가 큰 텍스트 영상에서는 두 클러스터가 잘 분류되지만 크기가 작은 영상에서는 텍스트와 배경의 경계부분이 뭉그러지는 현상이 발생하여 OCR 인식에 나쁜 영향을 주게 된다. 또한 텍스트에 해당하는 부분이 2개의 색상으로 구성된 경우에 텍스트의 일부가 삭제되는 단점이 존재하게 된다. (그림 3)은 2-means 클러스터링에 의해서 이진화가 잘 되지 않은 영상의 예를 보여준다.

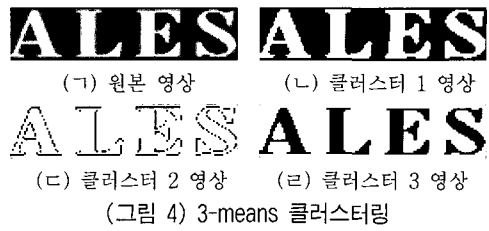
(그림 3)과 같은 단점을 보완하기 위해서 본 논문에서는 텍스트 후보 영상내의 클러스터를 다음과 같이 3종류(3-means)로 구분하였다.

- ① 배경 클러스터
- ② 텍스트 클러스터
- ③ 텍스트와 배경 경계면에 존재하는 클러스터

영상 내에 존재하는 클러스터를 3가지로 분류하였을 때 크기가 큰 영상과 작은 영상 모두에서 향상된 이진화 결과를 얻을 수 있으며, 텍스트 클러스터의 컬러가 두 개일 때 도 다른 텍스트 영역이 지워지지 않는다. 본 논문에서는 텍



(ㄱ) 원본 영상 (ㄴ) 클러스터링 결과  
(그림 3) 2-means 클러스터링의 오류



(그림 4) 3-means 클러스터링

스트 컬러가 2개 이내인 것으로 가정한다. 본 논문에서 사용한 3-means 클러스터링 알고리즘은 다음과 같다.

- 단계 1: 초기 3개 군집(cluster)의 중심을 임의로 선택한다.
- 단계 2: 주어진 중심점을 기준으로 각각의 픽셀을 가까운 군집에 할당한다. 이 때 중심점과 각 픽셀 간의 거리는 RGB 3차원 색상 공간에서의 유클리디안 거리로 계산한다.
- 단계 3: 새로 할당된 픽셀의 컬러 값을 중심으로 각 군집의 새로운 중심점을 계산한다.
- 단계 4: 만약 기존 중심점과 새로운 중심점간의 차이가 없으면 중지하고, 그렇지 않으면 단계 2로 돌아가 다시 수행한다.

(그림 4)는 입력영상에 3-means 클러스터링 알고리즘을 적용한 후, 각 클러스터에 해당하는 영상을 보여준다. 여기서 영상내의 흑화소가 관심 객체이다.

나. Text/Non-Text 클러스터 분류

대부분의 텍스트영상에 대해 3-means 클러스터링을 수행하고 나면 배경 클러스터, 텍스트 클러스터, 텍스트와 배경 경계면에 존재하는 클러스터로 군집화된다. 클러스터링된 3개의 클러스터 중에서 배경 클러스터는 다음과 같은 특징이 있다.

- ① 클러스터 영상내 흑화소 개수가 많다.
- ② 영상의 네 모서리를 포함한다.
- ③ 클러스터의 세로 높이가 영상의 세로 높이와 거의 같다.
- ④ 클러스터의 가로 넓이가 영상의 가로 넓이와 거의 같다.

위에서 서술한 4가지 속성을 만족하는 클러스터는 배경 클러스터로 결정한다. (그림 5)는 배경 클러스터가 가지고 있는 4가지 속성을 예시한다.

텍스트와 배경의 경계면에 존재하는 클러스터는 다음과 같은 특징이 있다.

- ① 클러스터 크기가 작다.
- ② 클러스터내의 흑화소 성분의 획의 두께가 작다.



1. 영상의 네 모서리를 포함  
2. 영상의 세로 높이  
3. 영상의 가로 넓이  
(그림 5) 배경 클러스터 속성

위의 속성 ②와 관련하여 식 (1)을 만족하면 경계면에 존재하는 클러스터로 결정하고, 그렇지 않으면 텍스트 클러스터로 간주한다.

$$\frac{\sum_{i=1}^n CC\_W_i}{n} < T_1 \quad \dots \quad \text{식 (1)}$$

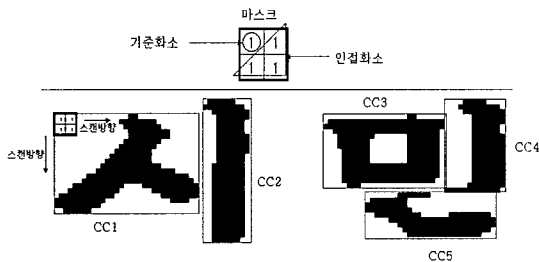
여기서  $CC\_W_i$ 는  $i$ 번째  $CC$ 의 두께이며,  $n$ 은 클러스터 내에 있는  $CC$ (Connected Component)의 개수,  $T_1$ 은 임계값이다. 획의 두께  $CC\_W_i$ 는 식 (2)와 같이 계산한다.

$$CC\_W_i = \frac{CC\_P\_S_i}{CC\_M\_N_i} \quad \dots \quad \text{식 (2)}$$

여기서  $CC\_P\_S_i$ 는 클러스터 내 흑화소 개수이고,  $CC\_M\_N_i$ 는 자신을 포함한  $2 \times 2$  이웃의 모든 화소가 흑화소인 경우의 수이다.

(그림 6)은 클러스터영상에서 획의 두께를 결정하기 위해 사용한 마스크와 영상 내에 존재하는 5개의  $CC$ 를 보여준다. 일반적으로 3-means 클러스터링 후에 흑화소를 가장 적게 포함하는 클러스터가 경계면 클러스터에 해당된다((그림 4) (㉔) 참조). 그러나 (그림 3)의 예에서와 같이 텍스트 색상이 두 가지인 경우, 가장 크기가 작은 클러스터라 할지라도 경계면이 아닌 텍스트 클러스터로 분류된다((그림 7) 참조).

3개의 클러스터 중에서 배경 클러스터와 경계면 클러스터를 제거하고 나면 (그림 8)과 같이 텍스트 클러스터 영역만 남게 된다. (그림 9)는 입력 영상에 2-means와 3-means 방법을 사용하여 이진화한 결과를 비교하고 있다. 실험 결과



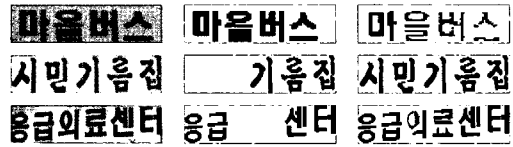
(그림 6) 획 두께를 결정짓는 마스크와 연결요소



(㉑) 원영상  
 (㉒) 가장 큰 클러스터(획두께=12)  
 (㉓) 중간 클러스터(획두께=7)  
 (㉔) 가장 작은 클러스터(획두께=7)  
 (그림 7) 경계면 클러스터 결정



(㉑) 원본 영상 (㉒) 이진화 영상  
 (그림 8) 이진화 영상



(㉑) 원본 영상  
 (㉒) 2-means 기반 이진화  
 (㉓) 3-means 기반 이진화  
 (그림 9) 2-means와 3-means 클러스터링 비교 결과

에서 알 수 있듯이 본 논문에서 제안한 3-means 알고리즘이 더 좋은 성능을 보여준다.

**다. 잡영 제거 및 이진화 성공 값 결정**

배경 클러스터와 경계면 클러스터를 제거하고 난 후에도 영상에는 잡영이 남아 있게 된다. 잡영 제거 및 이진화 성공 값을 결정하기 위해서 Blob Coloring[21] 방법을 적용하여 영상 내에 남아 있는 모든  $CC$ 를 추출하고, 이들 각각의 특성을 분석한다. 다음은 잡영제거 및 이진화 성공 값을 결정짓는 규칙이다.

- 단계 1:  $i$  번째  $CC$ 의 가로크기와 세로크기가 식 (3)을 만족하면 작은 잡영으로 판단 제거한다.

$$CC\_HLL_i < T_2 \text{ AND } CC\_WLL_i < T_3 \quad \dots \quad \text{식 (3)}$$

여기서  $CC\_HLL_i$ 는  $i$  번째  $CC$ 의 세로 크기이며,  $CC\_WLL_i$ 는  $i$  번째  $CC$ 의 가로 크기,  $T_2, T_3$ 는 임계값이다.

- 단계 2:  $i$  번째  $CC$ 의 가로크기와 세로크기가 식 (4)을 만족하면 텍스트  $CC$ 로 간주하고 다른  $CC$ 를 검사하기 위해 단계 1로 간다. 식 (4)을 만족하지 않으면 단계 3으로 간다.

$$CC\_HLL_i < T_4 \text{ AND } CC\_WLL_i < T_5 \quad \dots \quad \text{식 (4)}$$

여기서  $T_4, T_5$ 는 임계값이다. ( $T_4 > T_2, T_5 > T_3$ )

- 단계 3: 식(4)를 만족하지 않은  $i$  번째  $CC$ 의 최 외곽 상자(bounding box) 내에 있는 흑화소 수를 계산한 후에 식 (5)을 만족하면 이진화 실패값을 출력하고 이진화 Function을 종료한다. 식 (5)을 만족하지 않으면  $i$  번째  $CC$ 를 큰 잡영으로 판단하여 제거하고, 나머지  $CC$ 를 검사하기 위해 단계 1로 간다.



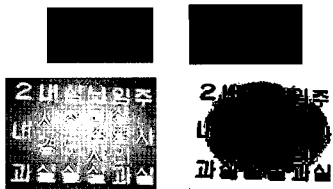
(ㄱ) 잡영 제거 전 영상 (ㄴ) 잡영 제거 후 영상  
(그림 10) 잡영 제거 영상

$$CC\_BLACK\_PSUM_i > T_6 \dots \quad \text{식 (5)}$$

여기서  $CC\_BLACK\_PSUM_i$ 는  $i$  번째 CC의 최 외곽상자 내의 흑 화소 수이다.

- 단계 4: 모든 CC의 처리가 종료되고 나면 이진화 성공 값을 결정하고 이진화 함수를 종료한다.

(그림 10)은 잡영 제거 및 이진화 성공 값 규칙을 적용한 후의 영상을 나타낸다. 식 3, 4, 5에서 사용한  $T_1, T_2, T_3, T_4, T_5, T_6$ 는 영상의 크기에 따라서 가변적인 값을 갖는다. 1단계의 이진화 과정을 거치고 나면 텍스트 컬러와 배경 컬러가 잘 구별되는 단순 영상은 이진화 성공 값이 True 값을 갖게 되어 이진화 과정을 멈추게 되고, 자연조명이나 인공조명에 의해서 영향을 받은 영상, 텍스트나 배경에 텍스처가 존재하는 영상은 이진화 성공 값이 False 값을 갖게 되어 다음 단계인 2단계 색상 분할을 수행하게 된다. (그림 11)은 1단계 색상 분할에 의해서 이진화가 잘 되지 않은 영상의 예를 보여준다.



(ㄱ) 원본 영상 (ㄴ) 이진화 된 영상

(그림 11) 1단계 색상 분할에 의해 이진화가 잘 안된 영상

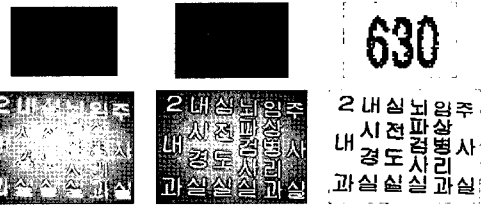
### 2.2 2단계 색상 분할

2단계 색상 분할은 입력 영상 중에서 자연조명이나 인공조명에 영향을 받은 영상의 이진화를 목표로 한다. 영상을 획득할 때 자연적이거나 인공적으로 첨가된 조명은 영상의 Contrast를 떨어뜨림으로서 이진화 결과에 나쁜 영향을 주게 된다. 제안하는 알고리즘에서는 고주파 통과 필터를 적용하여 영상의 세세한 부분의 Contrast를 증가시켜서 조명의 영향을 받은 영상의 향상된 이진화 결과를 얻을 수 있다. (그림 12)는 본 시스템에서 사용한 3×3 고주파 통과 필터를 나타낸다.

고주파 통과 필터를 통과한 영상은 1단계 색상 분할에서와 동일한 이진화 Function에 의해 이진화된다. 이때에도 이

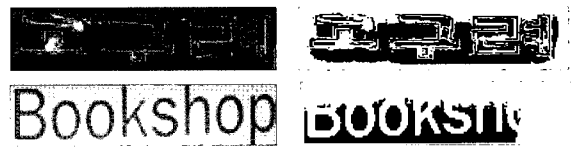
0	-1	0
-1	5	-1
0	-1	0

(그림 12) 3×3 고주파 통과 필터



(ㄱ) 원본 영상 (ㄴ) 고주파 통과 후 영상 (ㄷ) 이진화 후 영상

(그림 13) 2단계 색상 분할 후 영상



(ㄱ) 원본 영상 (ㄴ) 이진화 후 영상

(그림 14) 2단계 색상 분할에 의해 이진화가 잘 안된 영상

진화 성공 값이 True 값을 갖게 되면 이진화 과정을 멈추고, 이진화 성공 값이 False 이면 다음 단계인 3단계 색상 분할을 수행하게 된다. (그림 13은 2단계 색상 분할에 의한 이진화 결과를 보여준다. (그림 14)는 텍스트나 배경에 텍스처가 존재하는 영상으로써 2단계 색상 분할 후의 결과가 만족스럽지 않은 경우의 예를 보여준다.

### 2.3 3단계 색상 분할

3단계 색상 분할은 입력 영상 중에서 텍스트나 배경에 텍스처가 존재하는 영상의 이진화를 목표로 한다. 텍스트나 배경에 텍스처가 존재하는 영상은 영상의 세세한 부분 때문에 이진화 결과에 나쁜 영향을 주게 된다. 제안하는 알고리즘에서는 저주파 통과 필터를 적용하여 영상의 세세한 부분을 평활화함으로써 향상된 이진화 결과를 얻을 수 있다. (그림 15)는 본 시스템에서 사용한 3×3 저주파 통과 필터를 나타낸다.

$$\frac{1}{9} \times \begin{matrix} \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} \end{matrix}$$

(그림 15) 3×3 저주파 통과 필터



(ㄱ) 원본 영상 (ㄴ) 저주파 통과 후 영상 (ㄷ) 이진화 후 영상

(그림 16) 3단계 색상 분할 후 이진화된 영상들

저주파 통과 필터를 통과한 영상은 1단계 색상 분할에서 수행한 이진화 Function을 수행하게 된다. (그림 16)는 3단계 색상 분할에 의한 이진화 결과를 보여준다.

### 3. 실험 및 결과

본 논문의 실험에 사용한 영상들은 KAIST와 숙명여대가 공동 수집한 자연영상 120개와 2003 ICDAR Contest Image 487개 영상에 대해[18] 시스템을 적용하여 추출한 텍스트 후보 영상 510개이다. 영상내의 텍스트 종류는 한글, 영문, 한자, 숫자로 구성되어 있다. 입력 영상은 색상, 조명 및 텍스처에 따라서 텍스트 컬러와 배경 컬러가 잘 구별되는 Type 1 영상(410개), 자연조명이나 인공조명에 의해서 영향을 받은 Type 2 영상(65개) 및 텍스트나 배경에 텍스처가 존재하는 Type 3 영상(35개)으로 분류하였다.

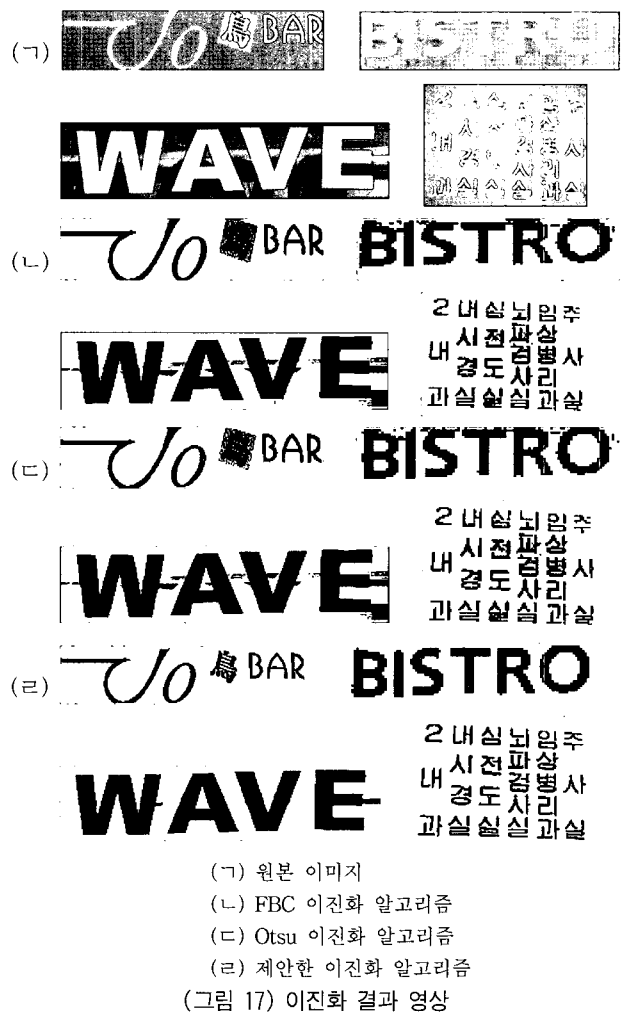
<표 4>와 <표 5>는 그레이 정보를 이용하는 이진화 알고리즘으로 많이 사용하고 있는 FBC 알고리즘[19]과 Otsu 알고리즘[20] 및 제안한 이진화 알고리즘을 이용하여 이진화한 후에 야르미 6.0을 사용하여 인식한 결과를 나타낸다. <표 4>, <표 5>에 있는 총 개수는 영상 내에 있는 전체 텍스트의 개수를 의미하고, 맞은 개수는 야르미 6.0을 사용하여 정확하게 인식한 텍스트의 개수를 의미한다. 제안 알고리즘의 성능은 필드단위로는 41.3(문자단위로는 57.6)%이며, 1단계만 실행할 경우 30.5(47.6)%, 1단계와 2단계를 실행할 경우 36.3(54.6)%이다. <표 4>, <표 5>의 실험 결과에서 알 수 있듯이 본 논문에서 제안한 3단계 색상 분할 이진화 알고리즘은 그레이 정보를 이용한 이진화 알고리즘인 FBC 알고리즘과 Otsu 알고리즘 보다 이진화 성능이 우수하다. (그림 17)은 추출된 텍스트 영상에 대해서 FBC 알고리즘, Otsu 알고리즘 및 제안한 알고리즘의 이진화 결과를 비교하여 보여준다.

<표 4> 필드단위 인식 결과

영상	알고리즘			
	FBC	Otsu	제안 알고리즘	
Type 1 영상	총개수	410	410	410
	맞은개수	22	25	171
	인식률(%)	5.0	6.0	41.7
Type 2 영상	총개수	65	65	65
	맞은개수	4	5	28
	인식률(%)	6.0	6.0	43.1
Type 3 영상	총개수	35	35	35
	맞은개수	0	0	12
	인식률(%)	0	0	34.3
합계	총합계	510	510	510
	맞은개수	26	30	211
	인식률(%)	5.0	6.0	41.3

<표 5> 문자단위 인식 결과

영상	알고리즘			
	FBC	Otsu	제안 알고리즘	
Type 1 영상	총개수	3,121	3,121	3,121
	맞은개수	571	580	1,967
	인식률(%)	18.2	18.5	63.0
Type 2 영상	총개수	687	687	687
	맞은개수	82	88	247
	인식률(%)	12.0	12.8	36.0
Type 3 영상	총개수	97	97	97
	맞은개수	10	12	37
	인식률(%)	10.0	12.3	38.2
합계	총합계	3,905	3,905	3,905
	맞은개수	663	680	2,251
	인식률(%)	17.0	17.4	57.6



### 4. 결론

본 논문에서는 자연 영상에서 추출된 텍스트 영상의 이진

화를 위한 3단계 색상 분할 알고리즘을 제안한다. 1단계 색상 분할은 입력 영상 중에서 텍스트색상과 배경색상이 잘 구별되는 단순 영상의 이진화를 수행하고, 2단계 색상 분할은 입력 영상에 고주파 필터를 적용하여 자연조명이나 인공조명에 영향을 받은 영상의 이진화를 수행하며, 3단계 색상 분할은 저주파 필터를 기반으로 텍스트나 배경에 텍스트가 존재하는 영상의 이진화를 수행한다. 제안한 이진화 알고리즘은 텍스트 컬러와 배경 컬러가 잘 구별되는 단순 영상뿐 아니라 자연조명이나 인공조명에 의해서 영향을 받은 영상 및 텍스트나 배경에 텍스트가 존재하는 영상 모두에서 그레이 정보를 이용한 이진화 알고리즘보다 효과적으로 텍스트 영역을 이진화 함을 보였다. 필드 및 문자단위의 인식 결과를 보면 제안 알고리즘의 성능은 필드단위로는 41.3(문자단위로는 57.6)%이며, FBC 알고리즘은 필드단위로는 5(문자단위로는 17%)이며, Otsu 알고리즘은 6(문자단위로는 17.4)%를 보여 제안한 이진화 알고리즘의 성능이 우수함을 알 수 있다. 이러한 장면 텍스트의 추출 및 인식에 대한 연구 결과는 향후 시각 장애인 보조 보조 시스템, 로봇 자동 주행 시스템, 비디오 자동 검색 및 색인 시스템, 텍스트 자동 번역 시스템, 영상과 함께 포함되어 전송되는 스팸 메일 필터 등과 같은 다양한 분야에서 널리 활용될 수 있을 것이다.

향후 연구 과제로는 텍스트 컬러가 3가지 이상인 영상 효율적으로 이진화하는 알고리즘 개발과 자연 영상 내에 존재하는 다양한 크기, 서체, 원근감을 갖는 텍스트를 인식할 수 있는 인식기의 개발이다.

## 참 고 문 헌

- [1] H.R. Byun, M.C. Roh, K.C. Kim, Y.W. Choi and S.W. Lee, "Scene Text Extraction in Complex Images," Proc. 5th International Workshop on Document Analysis Systems, pp.307-318, 2002.
- [2] D. Chen, H. Bourlard and H.P. Thiran, "Text Identification in Complex Background Using SVM," Proc. IEEE Computer Society Conference on CVPR, Vol.2, pp.621-626, 2001.
- [3] B.T. Chun, Y. Bae and T.Y. Kim, "Automatic Text Extraction in Digital Videos using FFT and Neural Network," Proc. IEEE Fuzzy Systems Conference, pp.1112-1115, Seoul, Korea, 1999.
- [4] P. Clark and M. Mirmehdi, "Combining Statistical Measures to Find Image Text Regions," Proc. 15th International Conference on Pattern Recognition, Vol.1, pp.450-453, 2000.
- [5] P. Clark and M. Mirmehdi, "Recognizing Text in Real Scene," International Journal of Document Analysis and Recognition, Vol.4, pp.243-257, 2002.
- [6] J. Gao and J. Yang, "An Adaptive Algorithm for Text Detection from Natural Scenes," Proc. IEEE Computer Society Conference on CVPR, Vol.2, pp.84-89, 2001.
- [7] O. Hori, "A Video Text Extraction Method for Character Recognition," Proc. Fifth International Conference on Document Analysis and Recognition, pp.25-28, 1999.
- [8] J. Ohya, A. Shio and S. Akamatsu, "Recognizing Characters in Scene Images," IEEE Trans. Pattern Analysis and Machine Intelligence, PAMI-16(2), pp.214-220, 1994.
- [9] X. Wang, X. Ding and C. Liu, "Character Extraction and Recognition in Natural Scene Images," Proc. Sixth International Conference on Document Analysis and Recognition, pp.1084-1088, 2001.
- [10] C. Wolf and J.M. Jolion, "Extraction and Recognition of Artificial Text in Multimedia Documents," Pattern Analysis and Applications, Vol.6, No.4, pp.306-326, 2003.
- [11] V. Wu, R. Manmatha and E.M. Riseman, "An Automatic System to Detect and Recognize Text in Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.21, No.11, pp.1224-1229, 1999.
- [12] J. Zhang, X. Chen, A. Hanneman, J. Yang and A. Waibel, "A Robust Approach for Recognition of Text Embedded in Natural Scenes," Proc. 16th International Conference on Pattern Recognition, Vol.3, pp.204-207, 2002.
- [13] 김길천, 최영우, 변혜란, "장면(Scene) 텍스트 추출 및 기울기/원근 추정", 제14회 영상처리 및 이해에 관한 워크샵 발표 논문집, pp.277-282, 제주도, 2002.
- [14] 김의정, 정원일, "칼라 문서에서 문자 영역 추출을 위한 클러스터링 기법", 대전산업대학교 논문집, 제14권, pp.107-116, 1997.
- [15] 김형권, 최원호, "자연 영상에서의 문자 패턴 추출", 울산대학교 공학연구논문집, 제26권 제2호, pp.35-54, 1995.
- [16] 노명철, 최영우, 이성환, "색상 및 명도 정보를 이용한 장면 텍스트 추출", 제14회 영상처리 및 이해에 관한 워크샵 발표 논문집, pp.515-520, 제주도, 2002.
- [17] 김지수, 김수형, "명도 정보를 이용한 자연 영상에서의 텍스트 영역 추출", 한국정보처리학회 호남·제주지부 학술발표논문집, 제3권 제1호, pp. 127-132, 2003.
- [18] 김지수, 김수형, 최영우, "명도 정보와 Split/Merge 분할을 이용한 자연 영상에서의 텍스트 영역 추출", 한국정보과학회 논문지 : 소프트웨어 및 응용, Vol.32, No.6, pp.502-511, 2005.
- [19] A. E. Savakis, "Adaptive Document Image Thresholding using Foreground and Background Clustering," Int. Conf. Image Proc. ICIP'98, Chicago, October, 1998.
- [20] N. Otsu, "A Threshold Selection Method from Gray-level Histograms," IEEE Trans. on System Man and Cybernetics, 9(1), pp.62-66, 1979.
- [21] D.H. Ballard and C.M. Brown, Computer Vision, Prentice-Hall, 1982.
- [22] 강나영, "시공간 데이터를 위한 클러스터링 기법의 성능 비교", 학위논문(석사), 이화여자대학교 과학기술대학원 : 컴퓨터학과, 2003. 8.



### 김 지 수

e-mail : kimjisoo@iip.chonnam.ac.kr

1996년 광주대학교 컴퓨터공학과(학사)

2003년 전남대학교 전산학과(이학석사)

2003년 3월~현재 전남대학교 전산학과  
박사과정

관심분야: 인공지능, 패턴인식, 문자인식



### 김 수 형

e-mail : shkim@chonnam.ac.kr

1986년 서울대학교 컴퓨터공학과(학사)

1988년 한국과학기술원 전산학과(공학석사)

1993년 한국과학기술원 전산학과(공학박사)

1993년~1996년 삼성전자 멀티미디어

연구소 선임연구원

1997년~현재 전남대학교 전자컴퓨터 정보통신공학부 부교수

관심분야: 인공지능, 패턴인식, 문서영상 정보검색, 유비쿼터스  
컴퓨팅