

# 카메라 획득 영상에서의 색 분산 및 개선된 K-means 색 병합을 이용한 텍스트 영역 추출 및 이진화

송 영 자<sup>†</sup> · 최 영 우<sup>\*\*</sup>

## 요 약

이미지에 포함된 텍스트는 이미지의 내용을 함축적이고 구체적으로 표현하는 정보로서 이러한 정보를 실시간에 찾아내서 인식한다면 다양한 응용에 활용할 수 있다. 본 논문에서는 카메라로 취득한 다양한 종류의 이미지로부터 텍스트를 추출하는 방법과 추출된 영역에서 텍스트를 분리하는 방법을 새롭게 제안한다. 텍스트 영역 추출을 위해서 RGB 색 공간에서 색 분산을 특징으로 제안하며, 텍스트 영역 분리를 위해서 RGB 색 공간에서 개선된 K-means 병합을 제안한다. 실험은 디지털 카메라와 핸드폰 카메라로 취득한 다양한 종류의 문서유형 이미지와 실내 외의 일반적인 자연이미지를 사용하였으며, ICDAR 콘테스트[1] 이미지의 일부도 사용하였다.

키워드 : 텍스트 특징, 색 분산, K-means 병합, 이진화

## Text Detection and Binarization using Color Variance and an Improved K-means Color Clustering in Camera-captured Images

Youngja Song<sup>†</sup> · Yeongwoo Choi<sup>\*\*</sup>

### ABSTRACT

Texts in images have significant and detailed information about the scenes, and if we can automatically detect and recognize those texts in real-time, it can be used in various applications. In this paper, we propose a new text detection method that can find texts from the various camera-captured images and propose a text segmentation method from the detected text regions. The detection method proposes color variance as a detection feature in RGB color space, and the segmentation method suggests an improved K-means color clustering in RGB color space. We have tested the proposed methods using various kinds of document style and natural scene images captured by digital cameras and mobile-phone camera, and we also tested the method with a portion of ICDAR[1] contest images.

Key Words : Text feature, Color Variance, K-means Clustering, Binarization

### 1. 서 론

최근 디지털 카메라와 휴대폰 카메라가 보편화 되면서 휴대용 기기에서의 이미지 획득이 간편하게 이루어지고 있다. 획득한 이미지에 포함되어 있는 문자를 빠르고 정확하게 추출하여 인식한다면 시각장애인의 보행 안내, 이동로봇의 주행 등과 같이 실생활에 유용하게 활용할 수 있다. 그 동안 이미지에서의 텍스트 추출 연구는 대부분 스캐너로 입력된 문서이미지나 자막이 포함된 동영상 이미지를 대상으로 수행되어 왔지만, 최근에는 디지털 카메라의 광범위한 보급으로 다양한 조명 및 복잡한 배경을 갖는 이미지에서의 텍스

트 영역 추출로 연구가 확장되고 있다.

기존의 연구로서 Jain 등은 이진, 웹, 색 이미지 및 비디오 프레임에서 텍스트를 추출하는 방법을 제안하였다[2]. 이진 및 웹 이미지에 대해서는 텍스트의 밝기 값이 균일하다는 특징을 사용하였고, 색 이미지 및 비디오 프레임에 대해서는 색 연속성을 특징으로 사용하였다. 제안한 방법은 서로 다른 종류의 이미지에 대해서 서로 다른 특징과 임계값을 적용해야하는 수동적인 실험을 수행하였으며, 이진 이미지, 웹 이미지 및 비디오 프레임에서는 높은 추출률을 보였지만, 색 이미지에 대해서는 낮은 추출률을 보였다. Kim 등은 비디오 프레임으로부터 텍스트 영역을 자동으로 추출하기 위해서 문자들이 수평 방향으로 위치하며 동일한 색과 일정한 크기를 갖는다는 가정을 기반으로 색 연속성 특징을 이용하였다[3]. 50개의 비디오 프레임에 제안한 알고리즘을

※ 본 연구는 숙명여자대학교 2005년도 교내연구비 지원에 의해 수행되었음.  
<sup>†</sup> 정 회 원 : 숙명여자대학교 컴퓨터과학과 박사과정  
<sup>\*\*</sup> 정 회 원 : 숙명여자대학교 컴퓨터과학과 교수  
 논문접수 : 2006년 1월 13일, 심사완료 : 2006년 5월 10일

적용한 결과 86%의 추출률을 보였으나, 색 대비가 크지 않은 텍스트와 크기가 작은 텍스트에 대해서는 오류가 발생하였다.

위와 같이 인위적인 제약조건을 갖는 텍스트 이미지와는 달리 다양한 배경과 크기의 텍스트를 갖는 자연이미지에서 텍스트 추출 연구로서 Ohya 등은 대부분의 텍스트들이 명도 값이 일정하다는 특징을 이용하였다[4]. 지역적 이진화 방법을 적용한 후 텍스트 영역 요소들의 유사성을 계산한 후 완화(relaxation) 방법을 적용하여 텍스트 영역 요소들을 서로 병합하였다. 텍스트 영역의 기울어짐이 없고 배경과 뚜렷하게 구분되는 단순한 이미지에서는 좋은 결과를 얻었지만, 복잡하거나 조명 변화가 있는 이미지에서는 오류가 많이 발생하였다. Ezaki 등은 시각장애인을 위한 자연이미지에서의 텍스트 추출을 연구하여, Sobel 에지 검출, Otsu 이진화 적용, 연결요소 추출 및 규칙기반 연결요소 필터링으로 구성된 텍스트 추출방법을 제안하였다[5]. 시각장애인을 위한 단순한 이미지에서의 추출은 가능했지만, 복잡한 배경과 색을 갖는 문자는 고려하지 않은 단점이 있다. 색을 이용한 텍스트 영역 추출 연구로서 Wang은 먼저 색 정보를 얻기 위해 다그룹분해(multi-group decomposition) 방식을 제안하였으며, 명도를 이용한 처리를 병행하였다[6]. 연결 요소와 잡영 제거, RLS(run length smearing) 과정을 거쳐 BAG(block adjacency graph) 알고리즘을 이용하였지만, 실험은 텍스트 색은 동일하며 텍스트 영역에는 어떤 그래픽도 포함하지 않는다는 전제를 두었다.

이미지에서 추출한 텍스트를 인식하기 위해서는 추출된 영역의 이진화가 선행되어야 하며, 이진화 방법은 전역적(global) 방법, 지역적(local) 또는 적응적(adaptive) 방법으로 구분된다[7]. 대표적인 전역적 방법으로서 Otsu가 제안한 방법이 있으며 비교적 단순한 배경을 갖는 이미지에 대해서는 좋은 성능을 보인다[8]. M. Seeger 등도 이미지 전체에 대한 전역적 임계치를 구하여 적용하는 이진화 방법을 제시하였지만 조명의 영향을 받는 단점이 있다[9]. 지역적 이진화 방법은 각 픽셀마다 또는 작은 영역마다 다른 임계값을 정함으로 하나의 이미지에 여러 개의 임계값이 설정된다. 김계경 등은 조명의 영향으로 텍스트 영역이 틀리게 분리되는 문제를 개선하기 위해서 국소 영역내의 최고 및 최소 명도 레벨의 차이로부터 임계값을 구하는 국소 적응적(local adaptive) 이진화 방법을 제안하였다[10]. 이 방법은 전반적으로 조명에 둔감하며 빠른 속도로 텍스트를 분리하나 색상만 다르지만 명도 차이가 작은 경우에는 배경과 텍스트를 분리하지 못하는 단점이 있다.

이미지에서 색 정보를 이용한 이진화 방법에 대한 연구도 진행되고 있다. Garcia 등은 비디오 프레임에서 문자 향상과 병합을 이용해서 문자 영역을 검출한 후, 색 양자화를 이용한 문자 분리 방법을 제안했다[11]. 이 과정에서 K-means 병합을 이용했으며 K는 항상 4로 고정된 후 HSI(hue saturation intensity) 색 공간에서 분리하였다. Wang 등은 색 이미지에서 텍스트를 이진화하기 위해서 색 줄임, 색 병합 및 문자 분리 과정을 수행했다[12]. RGB 색 공간에서 각 색 요소를 4비트로 줄인 후 그래프 이론 병합[13]을 이용해서

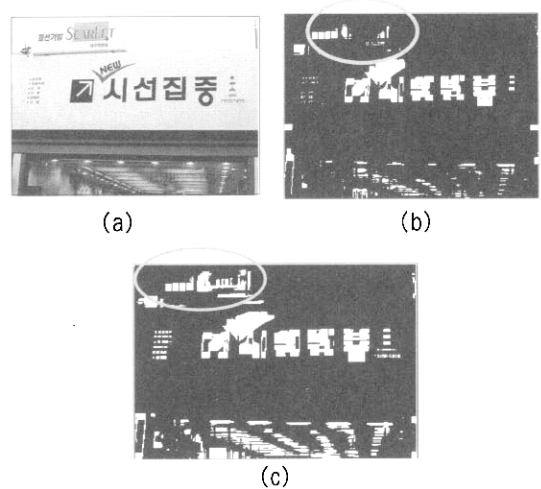
색 병합을 수행하였다.

본 논문에서는 텍스트 추출 특징으로서 RGB(red green blue) 색 공간에서의 색 분산을 텍스트 추출 특징으로 새롭게 제안한다. 제안한 특징은 배경과 문자영역의 색 차이가 작아서 명도이미지에서 구별하기 어려운 텍스트 영역들도 찾아내는 장점을 갖고 있다. 또한, 추출된 텍스트 영역으로부터 문자를 분리하기 위해서 RGB 색 공간에서의 색 병합을 제안하며, 이 방법은 초기 시드들과 K가 자동으로 선택 되도록 K-means 병합을 개선한 것이다.

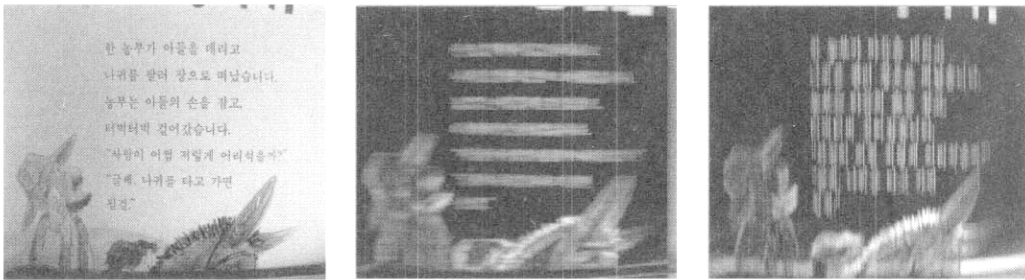
## 2. 텍스트 영역 추출

### 2.1 색 분산 특징

동영상에 포함된 자막이나 문서이미지에 포함된 텍스트는 배경과의 구별되는 색을 사용하거나 텍스트 주변을 단순하게 하여 텍스트와 배경의 구분을 쉽게 한다. 그러나 실내외에서 취득한 자연이미지는 빛과 조명 등의 영향으로 텍스트와 배경의 구분이 명확하지 않은 경우가 자주 발생한다. 본 논문에서는 이러한 이미지에서의 텍스트 추출의 정확성을 향상시키기 위해서 색 분산을 특징으로 제안한다. RGB 색 공간에서의 각 화소는 3차원으로 표현되어 3차원 공간에서의 분산을 측정하지만, 명도이미지에서의 분산은 1차원으로 표현되어 밝기 차이가 작은 이미지에서는 분산이 작게 표현된다. 이는 3차원 공간에서 서로 떨어져 있는 화소들이 1차원으로 투영되는 과정에서 거리가 짧아지기 때문이다. 색 사이의 거리를 정확하게 표현하기 위해서는 3차원 공간에서의 색 변화를 측정하는 것이 유리하다. (그림 1)은 주어진 이미지에 분산을 적용한 결과로서 (그림 1) (b)는 명도이미지에서의 분산이며, (그림 1) (c)는 RGB 색 이미지에서의 분산 이미지이다. 각 그림에서 타원으로 표시한 영역과 같이 글자와 배경의 명도 값이 비슷하여 구분이 어려운 영역도 색으로 구한 분산이미지에서는 보다 명확하게 구분되는 것을 확인할 수 있다.



(그림 1) 명도 및 색 분산 비교 : (a) 원 이미지, (b) 명도 분산 이미지, (c) 색 분산 이미지



(a) (b) (c)

(그림 2) 수평 및 수직 분산 이미지 : (a) 원 이미지, (b) 수평 분산 이미지, (c) 수직 분산 이미지

2.2 수평 및 수직 분산 이미지 생성

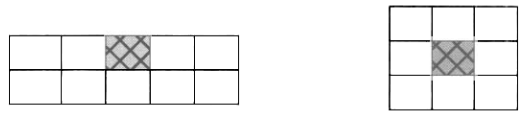
다양한 크기의 이미지를 320x240 크기로 만든 후 1x21 크기의 수평윈도우와 17x1 크기의 수직윈도우로서 식 (1)을 적용하여 분산을 구하며, 윈도우 크기는 실험으로 결정하였다. 식 (1)에서  $v_H$ 와  $v_V$ 는 수평윈도우 및 수직윈도우에서의 분산이며,  $R, G, B$ 는 픽셀의 색 값이다.  $\mu_{HR}, \mu_{HG}, \mu_{HB}$ 는 수평윈도우에서의  $R, G, B$  각각에 대한 평균이며,  $\mu_{VR}, \mu_{VG}, \mu_{VB}$ 는 수직윈도우에서의  $R, G, B$  각각에 대한 평균이다. 구해진 분산을 0과 255 사이의 값으로 재 분포시키며, (그림 2)는 수평 및 수직 방향으로 구해진 분산을 명도이미지로 표현한 것이다. 텍스트가 포함된 영역은 두 방향에서의 분산이 큰 것을 확인할 수 있다.

$$v_H = \frac{1}{21} \sum_{i=0}^{20} ((R_i - \mu_{HR})^2 + (G_i - \mu_{HG})^2 + (B_i - \mu_{HB})^2)$$

$$v_V = \frac{1}{17} \sum_{j=0}^{16} ((R_j - \mu_{VR})^2 + (G_j - \mu_{VG})^2 + (B_j - \mu_{VB})^2) \quad (1)$$

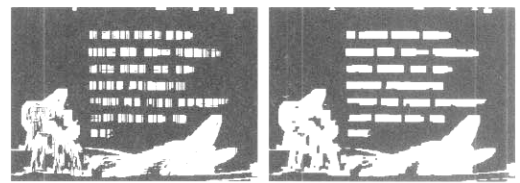
2.3 AND 연산을 이용한 잡영 제거와 영역 강조

텍스트 영역은 수평 및 수직 방향으로 분산이 모두 크게 나타나지만 가로나 세로방향의 선들은 한 방향에서만 분산이 크게 나타난다. 따라서 두 분산이미지에 대하여 논리적 AND를 수행하면 텍스트 주변의 불필요한 선들을 제거할 수 있다. AND 연산을 수행한 결과 이미지에는 분산은 크지만 텍스트가 아닌 영역들도 포함된다. 다음으로 분산 결과를 단순하게 만들고 잡영의 일부를 제거하기 위해서 모폴로지 연산을 적용한다. 모폴로지 연산은 수평 방향의 텍스트를 주로 강조하기 위해서 (그림 3)과 같은 2x5 크기의 불림(dilation) 연산자와 3x3 크기의 침식(erosion) 연산자를 구성하여 닫힘(closing)과 열림(opening) 연산을 반복 적용하였다. (그림 4) (a)는 수평, 수직 분산 이미지를 이진화한 후 논리적 AND 연산을 수행한 결과로서 이진화의 임계값은 분산이미지의 평균을 사용하였다. (그림 4) (b)는 AND 연산 결과에 제안한 모폴로지 연산자를 적용한 결과로서, 적용 전의 이미지로부터 부분적으로 잡영이 제거되고 텍스트 영역이 강조되어 연결요소의 수가 줄어든 이미지를 확인할 수 있다. (그림 4) (c)는 연결요소의 외곽사각형을 구한 결과이다.

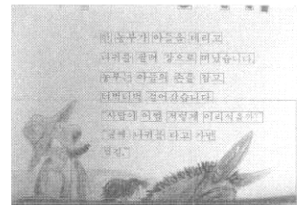


(a) (b)

(그림 3) 모폴로지 연산자 : (a) 불림 연산자, (b) 침식 연산자



(a) (b)



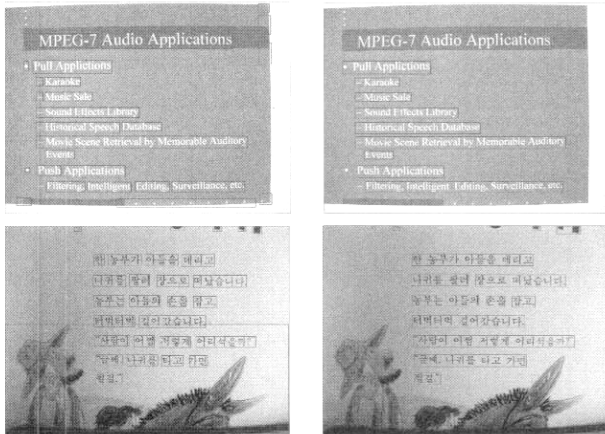
(c)

(그림 4) (a) AND 연산, (b) 모폴로지 적용, (c) 연결요소의 외곽사각형 추출

2.4 텍스트 영역 검증 및 추출

모폴로지 결과로부터 연결요소 외곽사각형의 위치, 크기, 종횡비(aspect ratio)와 연결요소의 크기를 후보영역 검증에 사용한다. 연결요소의 화소수가 아주 큰 경우는 에지 밀도가 조밀한 배경에서 발생된 경우가 대부분이며, 연결요소의 화소수가 아주 작은 경우는 비록 텍스트 영역이라 하더라도 사람의 눈으로 구분하기 어렵기 때문에 제외시킨다. 또한, 연결요소가 외곽사각형 안에서 차지하는 비율이 연결요소의 크기에 따라 적응적으로 변하는 임계값을 적용하여 텍스트 후보영역에서 제외한다. 외곽사각형의 폭 또는 높이가 아주 작은 요소들도 함께 제거한다.

추출된 텍스트 후보 영역들의 일부는 텍스트 열이 하나의 외곽사각형으로 존재하지만 일부는 글자 단위로 존재하기 때문에 단어 또는 문자 행 단위로 결합하는 것이 필요하다. 영역 결합을 위해서 외곽사각형의 포함정도, 거리 및 투영의 겹침 정도를 이용한다. 우선 외곽사각형의 포함 관계에



(그림 5) 추출된 텍스트 후보 영역 : 검증 및 결합 전, 후

서 큰 외곽사각형 안에 완전히 또는 대부분 포함되는 작은 외곽사각형은 큰 외곽사각형으로 대표한다. 다음은 두 외곽사각형의 X/Y축 방향으로 투영의 겹침 정도를 이용하여, 겹쳐지는 폭이 일정한 값 이상일 경우에 두 외곽사각형을 결합한다. 한 번 결합된 외곽사각형에 대해서는 다시 반복하여 외곽사각형의 X/Y축 방향으로 투영의 겹침 정도가 정해진 값을 넘는 외곽사각형이 없을 때까지 계속 반복한다. 끝으로 외곽사각형의 최단 거리를 구해서 서로 가까운 외곽사각형을 결합한다. (그림 5)는 영역 검출 및 검증을 거친 후 추출된 텍스트 후보 영역의 예를 보여준다. 적용한 겹침 정도는 83%, 외곽사각형과의 거리는 12 화소로 설정하였다.

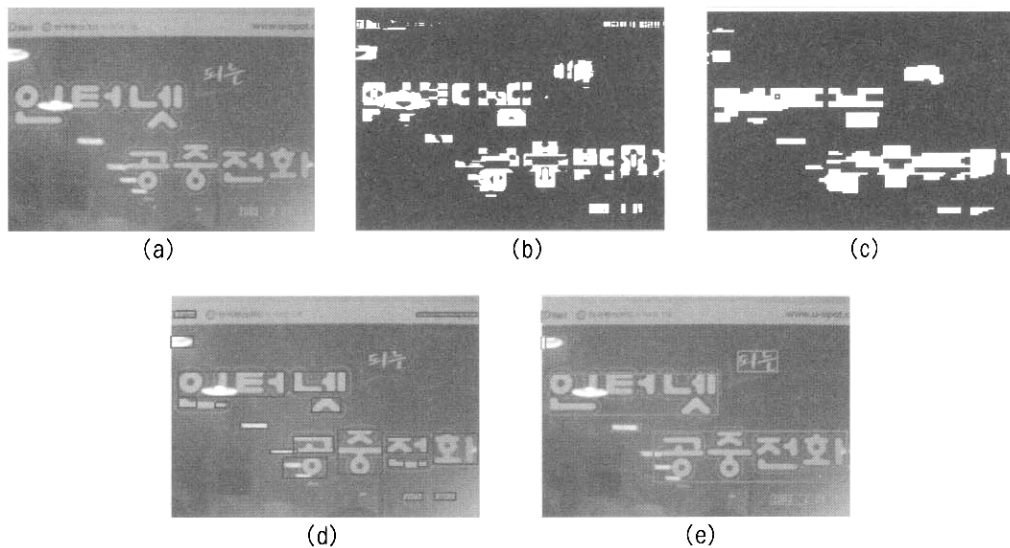
### 2.5 텍스트 크기를 고려한 영역 검출

입력된 이미지에서 글자의 크기가 아주 큰 경우에는 색 분산을 측정하는 윈도우가 글자의 폭과 거의 비슷하거나 작아져서 분산이 낮게 구해지는 경우가 발생한다. 이러한 오

류를 줄이기 위해서 320x240 크기의 이미지와 반으로 줄인 160x120 크기의 이미지를 동시에 사용하여 수평 및 수직 분산을 구한다. 이는 글자의 크기가 큰 경우 축소된 이미지에서 분산이 크게 나타나기 때문이다. 160x120 해상도에서는 1x15 크기의 수평방향 분산 윈도우와 11x1 크기의 수직방향 윈도우를 사용한다. 낮은 해상도의 이미지에서 두 분산 이미지를 AND연산한 후 3x3 크기의 구조연산자를 이용한 모폴로지 닫힘 연산을 수행하고, 검증 및 결합으로 텍스트 영역을 검출한다. 160x120의 낮은 해상도에서 구해진 외곽사각형 정보는 다시 320x240 해상도의 외곽사각형 크기로 환원시켜 두 해상도 모두에서 얻어진 외곽사각형의 포함, 겹침 정도, 최단 거리를 이용하여 최종 후보 영역을 결정한다. (그림 6)은 이러한 과정을 보여주며, (그림 6) (b)는 320x240 크기의 이미지에 대한 분산 및 모폴로지 결과이며, (그림 6) (c)는 160x120 크기의 이미지에 적용한 결과이다. (그림 6) (d)는 6(b)의 외곽사각형을 추출한 결과로서 윈도우 크기보다 작은 근접한 영역의 색 변화 값이 크지 않아서 외곽사각형의 결합에 어려움을 보여주고 있다. (그림 6) (e)는 6(c)의 연결요소 정보를 이용하여 외곽사각형을 구한 결과로서 고해상도 이미지에서 추출하지 못했던 영역들이 검출된 것을 보여준다.

### 3. 텍스트 분리

추출한 텍스트를 인식하기 위해서 텍스트를 배경과 분리하는 이진화 과정이 필요하다. 이미지 배경이 다양한 색으로 구성되거나 배경과 전경의 명도 차이가 작은 경우에는 기존의 이진화 방법으로 인식할만한 결과를 얻기가 어렵다. 본 논문에서는 이러한 문제점을 해결하기 위해서 색 병합을 이용한 이진화 방법을 제안한다. 제안하는 방법은 K-means 병합을 개선한 것이다.



(그림 6) 검출 결과 비교 : (a) 원이미지, (b) 320x240 이미지의 모폴로지 결과, (c) 160x120 이미지의 모폴로지 결과, (d) 320x240 이미지의 영역 검출 결과, (e) 160x120 이미지의 영역 검출 결과

### 3.1 색 병합

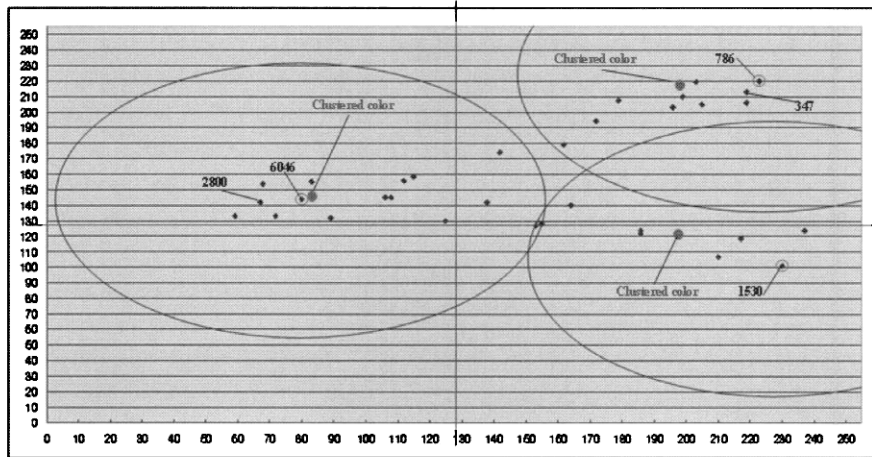
색 병합의 정확성을 향상시키기 위해서 320x240 크기의 축소된 이미지에서 검출한 텍스트 영역의 좌표를 축소하기 전의 원 이미지 좌표로 환원하여 원 이미지에서 텍스트 영역을 추출한다. 추출한 텍스트 영역은 다양한 크기를 가지며 각 화소는 24비트로 표현된 색 이미지이며, 여기에 개선된 K-means 병합을 수행한다. 개선된 K-means 방법은 K-means를 수행하기 전에 초기 평균(seed)과 클러스터(cluster)의 개수인 K가 자동으로 설정되도록 하였다. 먼저, RGB 색 공간에서 주어진 이미지의 색 히스토그램을 구성하고, 텍스트 영역 전체에 대해서 히스토그램이 0.5% 이하인 소수 색들은 병합에서 제외시킨다. 이 화소들은 병합이 끝난 후 가장 가까운 클러스터의 색 값으로 변경시킨다. 초기 시드 값은 색 히스토그램이 가장 큰 색으로 선택하는데 이는 가장 빈도가 높은 색이 텍스트 영역에서 중요한 색이기 때문이다. 다음은 초기에 선택된 색으로부터 일정한 색 거리 밖에 있는 색 중에서 히스토그램이 큰 색을 선택한다. 색 사이의 거리는 유클리디안 거리로 측정하며 색 공간에서 색들의 분포가 넓다고 판단되는 경우 크게 설정하고, 색들의 분포가 조밀하다고 판단되면 작게 설정하기 위해서 초기

시드와 다른 모든 픽셀들과의 거리의 합을 반영한다. 위의 과정을 반복하여 더 이상 일정한 거리 밖에 존재하는 색이 없을 때 선택된 시드들을 초기 시드로서 결정하며, 이 때의 시드 개수를 K로 설정한다.

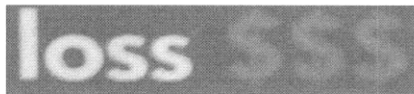
앞에서 구한 시드들과 K를 이용해서 K-means 병합을 수행한다. 입력 영상의 각 픽셀과 초기 평균과의 거리를 측정한 후 거리가 가까운 시드 영역으로 배정한 후 각 클러스터에서 새로운 평균을 구하는 과정을 반복하여 더 이상 평균의 변화가 없을 때 색 병합을 종료한다. (그림 7)은 K가 3인 경우로서 초기 시드 값과 K-means 병합 과정을 2차원 RG 평면으로 보인 것이며, 실제의 색 값들은 RGB 3차원 공간에서 결정된다. 각 숫자는 색 요소의 히스토그램을 나타내며, 원으로 둘러싸인 작은 점으로 표현한 색은 초기의 시드이며, K-means 방법이 적용된 후 이동한 시드의 위치는 큰 점으로 표현했다. (그림 8)은 개선된 K-means 방법을 적용하여 얻어진 색 병합 결과로서 텍스트 분리에 유리한 것을 확인할 수 있다.

### 3.2 색 면 선택

제안한 색 병합 방법을 적용하면 대부분 2개에서 5개 이



Seed1: (80, 144, 99)    Seed2: (223, 220, 209)    Seed3: (230, 101, 91)  
 병합 후: (82, 142, 99)    병합 후: (206, 211, 198)    병합 후: (198, 121, 74)

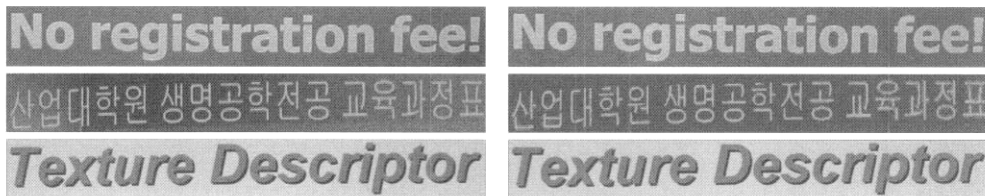


텍스트 영역 이미지

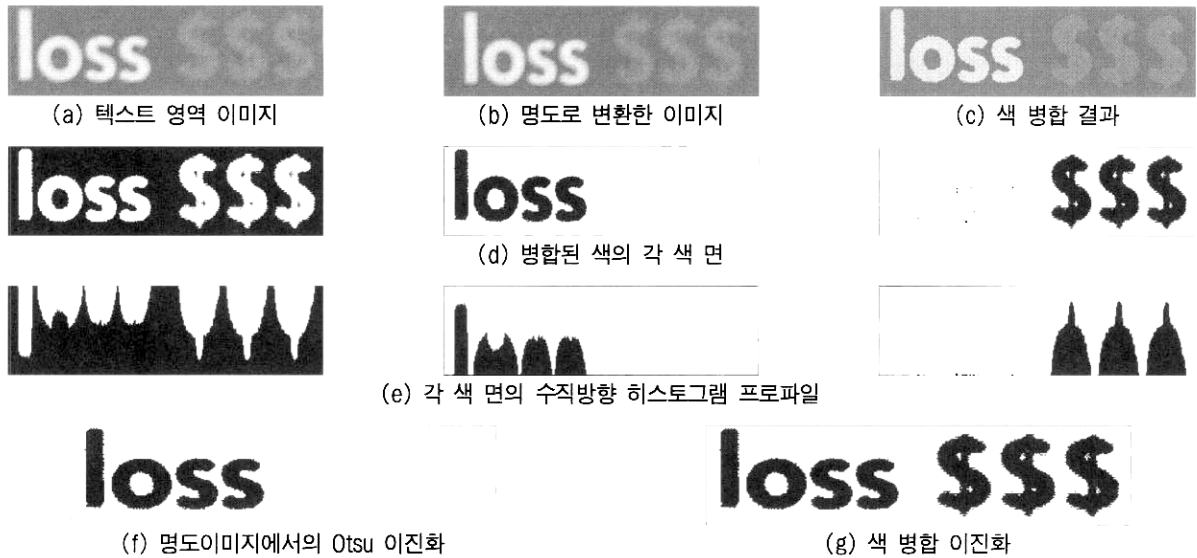


색 병합 결과

(그림 7) K가 3으로 설정된 경우의 초기 시드와 색 병합



(그림 8) 제안한 방법을 이용한 색 병합 결과



(그림 9) Otsu 이진화 및 색 병합을 이용한 이진화 결과 비교

내의 색으로 줄어든다. 병합된 색들은 각각의 색 면으로 분리한 후, 각 색 면들의 수직 방향 히스토그램 프로파일을 분석한다. 색 면들은 텍스트, 배경, 잡영 및 텍스트 테두리로 분류한다. (그림 9)는 색 병합된 후 각 색 면의 수직 히스토그램 프로파일을 보여준다. 텍스트를 포함하는 색 면은 배경에 비해서 프로파일의 히스토그램이 작고 히스토그램이 0인 값이 존재하며, 영역의 높이와 같은 히스토그램이 존재하지 않는 특징이 나타난다. 또한, 각 빈의 평균을 기준으로 0에서 1로 변하는 빈도수가 배경에 비해서 높게 나타나는 특징도 있다. 이러한 특징을 갖는 색 면들을 선택하여 이진화 결과로 얻을 수 있다. (그림 9)는 색 면 선택에 사용된 각각의 색 면에 대한 수직방향 히스토그램 프로파일, 명도이미지에서의 이진화 결과와 색 병합을 이용한 이진화 결과를 보여준다. 각 색 면의 수직방향 히스토그램에서 왼쪽 히스토그램은 값이 0이 아닌 빈이 존재하지 않지만, 영역의 높이와 같은 빈의 개수가 많기 때문에 배경으로 분류한다. 나머지 두 개의 색 면은 위의 텍스트 영역 특징을 만족하기 때문에 텍스트 영역으로 판단한다. 명도이미지에서의 이진화 결과와 비교하면 텍스트를 정확하게 분리한 것을 보여준다.

#### 4. 실험 및 결과

본 연구에서 제안한 방법을 윈도우즈 XP에서 Visual C++ 6.0을 이용하여 펜티엄 IV 1.8GHz 하드웨어에서 구현하였다. 제안한 텍스트 영역 검출 방법을 평가하기 위해서 두 종류의 디지털 카메라와 한 종류의 핸드폰 카메라로 취득한 이미지를 사용하였다. 실험 이미지는 카메라로 취득한 문서유형 이미지와 실내외 자연이미지로 나누었으며, 문서유형 이미지에는 동화책, 명함, 발표 자료 이미지를 포함하고, 실내외에서 취득한 이미지에는 간판, 표시판, 광고, 건물내의 표시판 등을 포함한다. 또한, ICDAR(International

Conference on Document Analysis and Recognition) 콘테스트에서 공개한 학습용 자연이미지의 일부도 포함하여 실험하였다.

평가는 식 (2)를 이용한 정확률(precision)과 재현율(recall)을 측정하였다[14]. 정확률은 텍스트로 검출한 것 중에서 정확한 텍스트의 비율을 표현하며, 이는 비텍스트가 텍스트로 분류되는 오류가 많아져도 정확률은 낮아진다. 재현율은 비텍스트가 텍스트로 분류된 것은 상관없이 단지 텍스트가 몇 개 찾아졌는가를 평가하는 것이다. 따라서 응용에 따라 정확률 또는 재현율에 우선권을 둘 수 있다. 본 논문에서는 제안한 방법은 높은 재현율을 보여주는데, 이는 텍스트 영역을 찾지 못하는 경우는 거의 발생하지 않음을 보여준다.

$$\begin{aligned} \text{정확률(precision)} &= (\text{정확히 검출한 텍스트 개수}) / (\text{검출한 텍스트 개수}) \\ \text{재현율(recall)} &= (\text{정확히 검출한 텍스트 개수}) / (\text{전체 텍스트 개수}) \end{aligned} \quad (2)$$

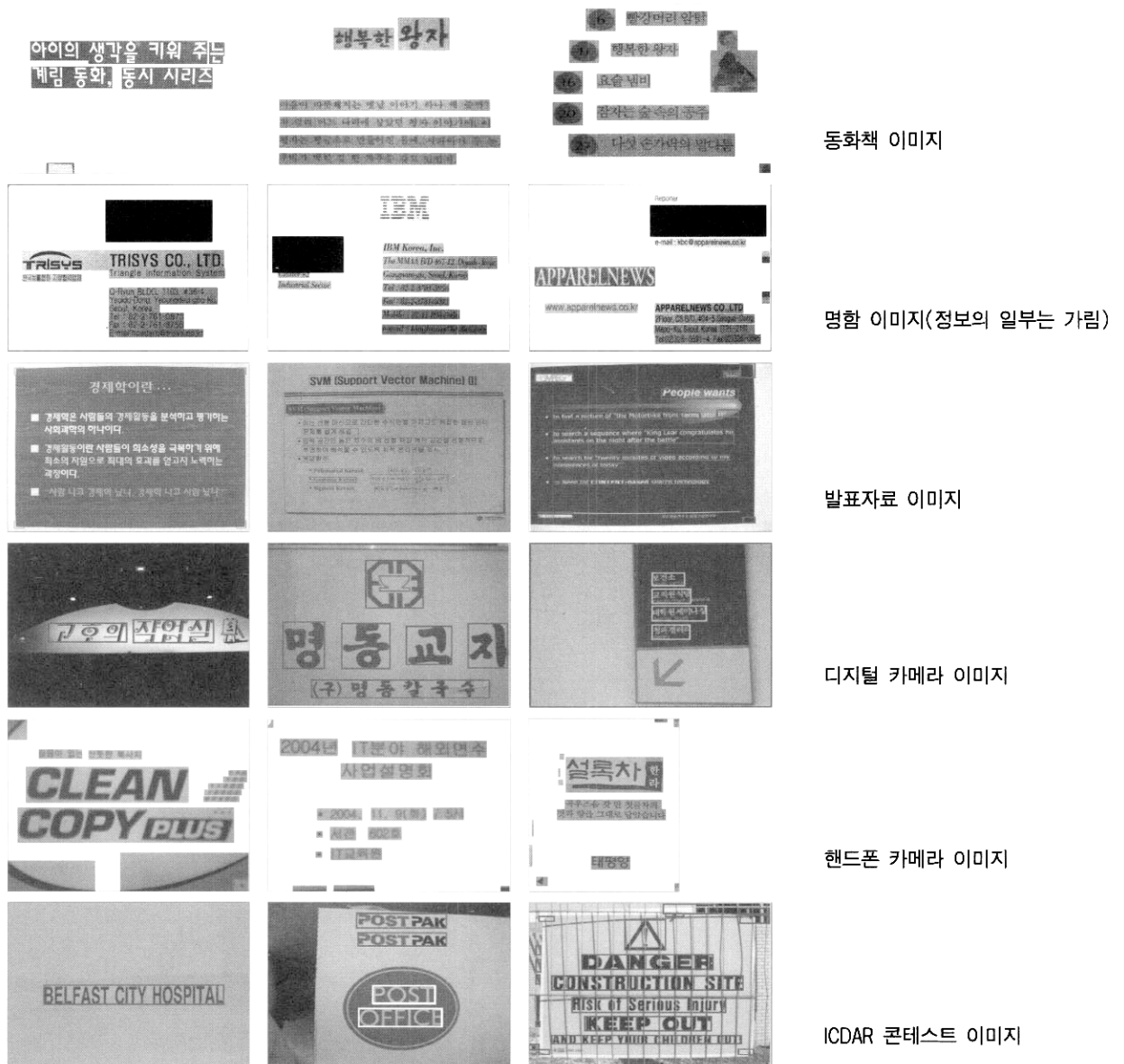
<표 1>과 <표 2>는 텍스트 영역 추출 결과를 보여주며 문서유형의 이미지에서 틀리게 검출한 텍스트 개수가 일반적인 자연이미지에 비하여 아주 적었다. 문서유형 이미지는 비교적 배경이 단순하며 글자의 크기나 폰트가 일정하지만, 일반적인 자연이미지는 복잡한 배경을 갖거나 동일한 이미지 내에서도 텍스트의 크기 차이가 심한 경우가 많다. 전체적으로 색 분산을 특징으로 사용하여 텍스트 영역 검출의 재현율이 95% 이상으로 대부분의 영역을 검출된 것을 확인할 수 있었다. (그림 10)은 각 실험 이미지의 유형별 검출 결과를 보여주며 카메라의 종류나 입력된 이미지의 해상도에 별다른 영향을 받지 않고 검출된 것을 확인할 수 있다. (그림 10)은 검출된 영역만을 보인 이미지와 전체 이미지에서 추출된 영역만을 표시한 이미지로 나누어 표현했다. 또한 최근의 텍스트 영역 추출 결과로서 2005년 ICDAR 콘테

〈표 1〉 문서유형 이미지에서의 텍스트 영역 추출 결과

		전체 텍스트 개수	맞게 검출한 개수	못 찾은 개수	틀리게 검출한 개수	정확률	재현률
동화책 이미지(100개)		1245	1230	15	257	0.83	0.99
명함 이미지(66개)		704	703	1	19	0.97	0.99
발표자료이미지	소니카메라 이용 (100개)	1037	1001	36	237	0.81	0.97
	코닥카메라 이용 (90개)	918	880	38	149	0.86	0.96

〈표 2〉 실내외 일반 이미지에서의 텍스트 영역 추출 결과

		전체 텍스트 영역 개수	맞게 검출한 개수	못 찾은 개수	틀리게 검출한 개수	정확률	재현률
디지털 카메라	소니(100개)	312	306	6	98	0.76	0.98
	코닥(100개)	538	523	15	127	0.80	0.97
핸드폰 카메라(100개)		601	587	14	101	0.85	0.98
ICDAR 이미지(공개된 이미지 일부 70개)		285	283	2	80	0.79	0.99



(그림 10) 다양한 이미지에서의 추출 결과



색 병합 결과



배경1과 히스토그램 프로파일



배경2와 히스토그램 프로파일



텍스트와 히스토그램 프로파일



색 병합 결과



배경과 히스토그램 프로파일



텍스트1과 히스토그램 프로파일



텍스트2와 히스토그램 프로파일

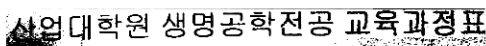
(a) 두 개의 배경과 하나의 텍스트로 이루어진 경우

(b) 하나의 배경과 두 개의 텍스트로 이루어진 경우

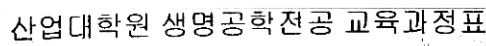
(그림 11) 색 면 선택 과정



(a)



(b)



(c)



(그림 12) 이진화 결과 비교 : (a) 원 이미지, (b) Otsu 이진화, (c) 제안한 방법의 결과

스트에서는 501개의 비공개된 테스트 이미지에 대해서 60%의 정확률과 67%의 재현률이 가장 우수한 결과로 보고 되었다[15]. 이는 공개된 학습 이미지와 공개되지 않은 테스트 이미지의 수준을 추측할 수 있는 수치로 판단되며, 본 논문에서 제안한 방법의 정확한 평가를 위해서는 2007년 ICDAR 콘테스트에 참가하여 확인하는 것이다.

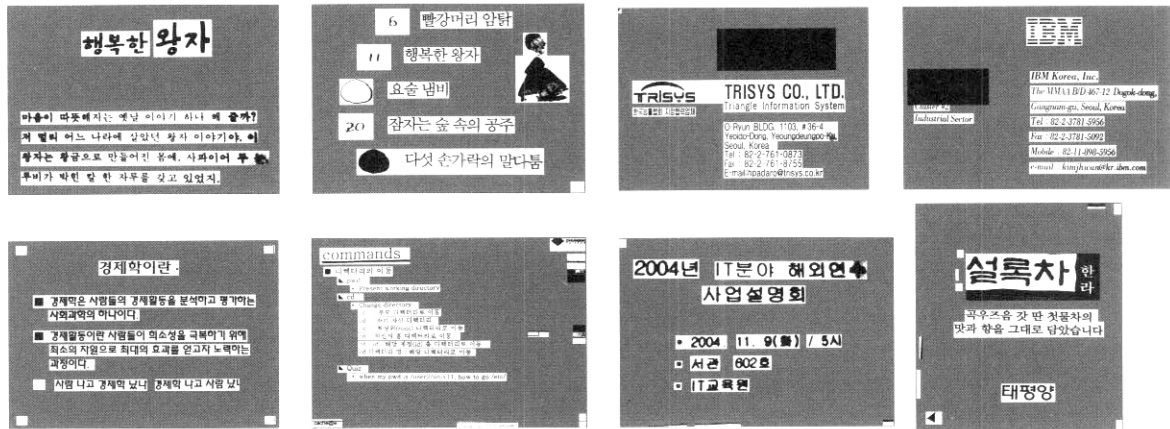
텍스트 분리 방법은 검출된 텍스트 영역을 기본 단위로 하여 인식할 수 있는 수준으로 문자가 이진화 되었는지를 사용자가 정성적으로 판단하였다. (그림 11)은 색 병합 결과에 따른 색 면 분리 과정으로서 각 색 면에 대하여 히스토그램 프로파일을 구한 후 배경과 텍스트를 분류하였다. 각

영역에서 텍스트의 개수와 배경의 개수에 따른 결과를 보여 준다.

(그림 12)는 배경 색이 점차 변하는 복잡한 색 분포 이미지와 조명의 영향으로 배경색이 점차 변하는 이미지에 대하여 대표적인 이진화 방법인 Otsu 방법과 제안한 방법의 결과를 비교한 것이다. 제안한 방법의 향상된 결과를 확인할 수 있다. (그림 13)은 (그림 10)의 일부 이미지에 대한 이진화 결과로서 명함이미지의 개인적인 정보 일부는 가렸다.

실험 이미지에 대한 처리 시간은 텍스트 영역 추출과 추출된 영역의 이진화 시간의 합으로서, 보통 단순한 배경을 갖는 이미지에 대해서는 평균 0.98초, 복잡한 배경을 갖는





(그림 13) 제안한 방법의 이진화 결과

이미지에 대해서는 평균 1.87초가 소요되었다.

### 5. 결 론

본 논문에서는 색 분산을 이용하여 텍스트 영역을 추출하는 방법과 추출된 텍스트 영역에서 텍스트를 이진화 시키는 방법을 새롭게 제안하였다. 제안한 분리 방법은 초기 시드와  $K$ 를 자동으로 결정하는 장점이 있으며, 적용한 결과 향상된 분리 결과를 확인할 수 있었다. 다양한 종류의 자연 이미지에 제안한 방법을 적용한 결과 높은 재현율을 얻어서 응용시스템에 적용할 수 있는 가능성을 확인하였다. 향후 연구로는 텍스트 영역 추출의 정확성을 향상시키기 위해서 각 과정별로 개선하는 것과 추출된 영역에서의 이진화 결과를 상용 OCR로 인식률을 측정해서 다른 방법들과의 비교를 객관적으로 수행하는 것이다.

### 참 고 문 헌

[1] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong and R. Young, "ICDAR 2003 Robust Reading Competition," *Proceeding of International Conference on Document Analysis and Recognition*, Vol.2, pp.682-687, 2003.

[2] Anil K. Jain and Bin Yu, "Automatic Text Location in Images and Video Frames," *Pattern Recognition*, Vol.31, No.12, pp.2055-2076, 1998.

[3] H. K. Kim, "Efficient Automatic Text Location Method and Content-based Indexing and Structuring of Video Database," *Journal of Visual Communications and Image Representation*, Vol.7, pp.336-344, 1996.

[4] J. Ohya, A. Shio and S. Akamatsu, "Recognizing Characters in Scene images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-16(2), pp.67-82, 1994.

[5] N. Ezaki, M. Bulacu and L.Schomaker, "Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons," *Proceedings of 17th International Conference on Pattern Recognition*, Vol.II, pp.683-686, 2004.

[6] Hao Wang, "Automatic Character Location and Segmentation in Color Scene Images," *Proceedings of 11th International Conference on Image Analysis and Processing*, pp.2-7, 2001.

[7] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, Addison Wesley, 1993.

[8] N. Otsu, "A thresholding selection method from gray-level histogram," *IEEE Transactions on System, Man, and Cybernetics*, No.9, pp.62-66, 1979.

[9] M. Seeger and C. Dance, "Binarizing camera images for OCR," *Proceeding of International Conference on Document Analysis and Recognition*, Vol.1, pp.54-58, 2001.

[10] 김계경, 지수영, 정연구, 박상규, "조명이 적은 카메라기반 문서 영상 인식시스템," *컴퓨터비전 및 패턴인식 연구회 워크샵*, pp.90-92, 2002.

[11] C. Garcia and X. Apostolidis, "Text detection and segmentation in complex color images," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol.4, pp.2326-2329, 2000.

[12] B. Wang, X-F. Li, F. Liu and F-Q. Hu, "Color text image binarization based on binary texture analysis," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol.3, pp.585-588, 2004.

[13] J. Matas and J. Kittler, "Spatial and feature space clustering: Applications in image analysis," *Proceedings of 6th International Conference on Computer Analysis of Images and Patterns*, pp.162-173, 1995.

[14] M. Junker and R. Hoch, "On the Evaluation of Document Analysis Components by Recall, Precision, and Accuracy," *Proceeding of International Conference on Document Analysis and Recognition*, pp.713-716, 1999.

[15] Simon M. Lucas, "ICDAR 2005 Text Locating Competition Results," *Proceeding of International Conference on Document Analysis and Recognition*, pp.80-84, 2005.



### 송 영 자

e-mail : soyo303@sookmyung.ac.kr

1989년 경북대학교 수학과(이학사)

1991년 경북대학교 컴퓨터공학과  
(공학석사)

2001년~2006년 숙명여자대학교  
컴퓨터과학과(이학박사)

관심분야: 영상처리, 문자추출, 패턴인식



### 최 영 우

e-mail : ywchoi@sookmyung.ac.kr

1985년 연세대학교 전자공학과(학사)

1986년 University of Southern California  
컴퓨터공학과(석사)

1994년 University of Southern California  
컴퓨터공학과(박사)

1994년~1997년 LG전자기술원 선임연구원

1997년~현재 숙명여자대학교 컴퓨터과학과 교수

관심분야: 영상처리, 패턴인식, 문자인식 등