

형태학과 문자의 모양을 이용한 뉴스 비디오에서의 자동 문자 추출

(Automatic Text Extraction from News Video using Morphology and Text Shape)

장 인 영 [†] 고 병 철 ^{**} 김 길 천 ^{**} 변 혜 란 ^{***}
 (InYoung Jang) (ByoungChul Ko) (KilCheon Kim) (HyeRan Byun)

요 약 최근 들어 인터넷 사용의 증가와 더불어 디지털 비디오의 수요 또한 급격히 증가하고 있는 추세이다. 따라서 디지털 비디오 데이터베이스의 인덱싱을 위한 자동화된 도구가 필요하게 되었다. 디지털 비디오 영상에 인위적으로 삽입되어진 문자와 배경에 자연적으로 포함되어진 배경문자 등의 문자 정보는 이러한 비디오 인덱싱을 위한 중요한 단서가 되어질 수 있다. 본 논문에서는 뉴스 비디오의 정지 영상에서 뉴스 자막과 배경 문자를 추출하기 위한 새로운 방법을 제안한다. 제안된 알고리즘은 다음과 같이 세 단계로 구성된다. 첫 번째 전처리 단계에서는 입력된 컬러 영상을 명도 영상으로 변환하고, 히스토그램 스트레칭을 적용하여 영상의 수준을 향상시킨다. 이 영상에 적응적 임계값 추출에 의한 분할 방법을 수정 적용하여 영상을 분할한다. 두 번째 단계에서는 적응적 이진화가 적용된 결과 영상에 모폴로지 연산을 적절하게 사용하여, 우선 문자 영역은 아니면서 문자로 판단되기 쉬운 양의 오류(false-positive) 요소들이 강조되어 남아있는 영상을 만든다. 또한, 변형된 이진화 결과 영상에 모폴로지 연산과 본 논문에서 제안한 기하학적 보정(Geo-correction) 필터링 방법을 적용하여 문자와 문자로 판단되기 쉬운 요소들이 모두 강조되어 남아있는 영상을 만든다. 이 두 영상의 차를 구함으로써 찾고자 하는 문자 요소들이 주로 남고, 문자가 아닌 문자처럼 보이는 오류 요소들은 대부분 제거된 결과 영상을 만든다. 문자로 판단되는 양의 오류 영역들을 남기는데 사용된 모폴로지 연산은 3x3 크기의 구조 요소를 갖는 열림과 (열림닫힘+닫힘열림)/2 이며, 문자 및 문자와 유사한 요소들을 남기는데 사용된 연산은 (열림닫힘+닫힘열림)/2와 기하학적 보정이다. 세 번째 검증 단계에서는 전체 영상 화소수 대비 각 후보 문자 영역의 화소수 비율, 각 후보 문자 영역의 전체 화소수 대비 외곽선의 화소수 비율, 각 외곽 사각형의 폭 대 높이의 비율 등을 고려하여 비문자로 판단되는 요소들을 제거한다. 임의의 300개의 국내 뉴스 영상을 대상으로 실험한 결과 93.6%의 문자 추출률을 얻을 수 있었다. 또한, 본 논문에서 제안한 방법으로 국외 뉴스, 영화 비디오 등의 영상에서도 좋은 추출을 보임을 확인할 수 있었다.

키워드 : 문자추출, 비디오 인덱싱, 모폴로지

Abstract In recent years the amount of digital video used has risen dramatically to keep pace with the increasing use of the Internet and consequently an automated method is needed for indexing digital video databases. Textual information, both superimposed and embedded scene texts, appearing in a digital video can be a crucial clue for helping the video indexing. In this paper, a new method is presented to extract both superimposed and embedded scene texts in a freeze-frame of news video. The algorithm is summarized in the following three steps. For the first step, a color image is converted into a gray-level image and applies contrast stretching to enhance the contrast of the input image. Then, a modified local adaptive thresholding is applied to the contrast-stretched image. The second step is divided into three processes: eliminating text-like components by applying erosion, dilation, and (OpenClose+CloseOpen)/2 morphological operations, maintaining text components using (OpenClose+

· 본 연구는 한국과학재단 목적기초연구(R04-2001-000-00075-0) 지원

으로 수행되었음

† 학생회원 : 연세대학교 컴퓨터학과
 stefano@aipiri.yonsei.ac.kr

** 비 회 원 : 연세대학교 컴퓨터학과
 soccer1@aipiri.yonsei.ac.kr

kimkch@aipiri.yonsei.ac.kr

*** 종신회원 : 연세대학교 컴퓨터과학과 교수

hrbyun@aipiri.yonsei.ac.kr

논문접수 : 2001년 10월 22일

심사완료 : 2002년 2월 19일

CloseOpen)/2 operation with a new Geo-correction method, and subtracting two result images for eliminating false-positive components further. In the third filtering step, the characteristics of each component such as the ratio of the number of pixels in each candidate component to the number of its boundary pixels and the ratio of the minor to the major axis of each bounding box are used. Acceptable results have been obtained using the proposed method on 300 news images with a recognition rate of 93.6%. Also, my method indicates a good performance on all the various kinds of images by adjusting the size of the structuring element.

Key words : Text extraction, Video indexing, Morphology

1. 서론

최근 들어 인터넷 사용의 증가와 더불어 디지털 비디오의 수요 또한 급격히 증가하고 있는 추세이다. 따라서 디지털 비디오 데이터베이스의 인덱싱을 위한 자동화된 도구가 필요하게 되었고 이에 따라 다양한 방법이 연구되어 오고있다. 기존의 전형적인 디지털 비디오 인덱싱 방법은 관리자가 비디오를 보고 직접 분할한 후 적절한 색인어를 입력하는 방식으로 이는 상대적으로 방대한 양의 디지털 비디오를 모두 처리하기에는 상당한 시간과 인력을 필요로 하기 때문에 극히 비효율적이며 또한 관리자 개개인의 주관에 의해 다른 색인어를 입력할 수 있으므로 색인어 입력에 있어서 오류를 범할 소지가 많다[1]. 현재까지 비디오 자동 인덱싱을 실현하기 위해 샷 분할 방법과 문자 추출 방법이 제안되어져 왔다. 샷 분할 방법은 그 자체만으로는 디지털 비디오 인덱싱의 실현을 위해 한계가 있기 때문에 최근에는 비디오 문자 추출 분야로 관심이 모아지고 있는 상황이며 디지털 비디오에 나타나는 자막문자와 배경에 포함되어진 문자 정보들은 이러한 문제를 해결할 수 있는 중요한 단서가 된다[1-5, 7]. 또한 최근의 논문들에서 이러한 문제를 해결하기 위한 다양한 해결책들을 찾을 수 있다. 일반적으로 이 분야에서 전형적인 문제점은 문자들의 다양한 크기와 서체, 문자들의 방향과 위치의 변화, 다른 질감, 일정치 않은 빛의 변화, 불규칙한 배경, 그리고 문자 획의 컬러 gradient등이다[6, 7, 14]. 본 논문에서는 문자 추출의 이러한 문제들을 해결하기 위해 모폴로지 연산과 기하학적 보정 방법을 선 처리 과정으로 제안하고 후보 자막 영역 중 실제 자막 영역만을 추출하기 위해 필터링 방법을 적용한다.

2. 관련연구

영상들에 존재하는 문자들을 자동으로 추출하고 인식하기 위한 다양한 접근법들을 최근의 논문에서 찾아볼 수 있다. 일반적으로 서적이거나 신문과 같은 이진화 영상

에 존재하는 문자 영역의 추출은 전경(foreground)의 화소수를 수평방향으로 계산하여 그 화소수가 '0'이면 공백 영역으로 그 이상이면 문자 영역으로 판단함으로써 문자를 구별해낼 수 있다[8]. 그러나 이 같은 방법은 일반적인 비디오 영상이나 풍경 영상에서 복잡한 질감(textual)과 그림요소(graphical objects)들 사이에 존재하는 문자들을 인식하고 추출하기 위한 도구로는 부족하다. 따라서 이러한 복잡한 영상에서 문자들을 추출해내기 위한 다양한 방법들이 제안되어져 왔는데 본 장에서는 이와 관련된 논문들과 관련 연구를 간단히 살펴본다. 먼저 Ohya *et al.*[9] 은 실외 영상에서 문자 추출과 인식에 대한 방법을 제안하고 있다. 이 논문에서는 문자들에 대해 몇 가지의 가정을 하고 있는데 첫째는 문자들의 방향이 위로 향해 있어야 한다는 것과 둘째, 배경과 분리되어져 있어야 한다는 것 그리고 마지막으로 일정한 그레이 색상을 가지고 있어야 한다는 것이다. 알고리즘은 먼저 적응적 임계값을 적용하여 후보 문자들의 패턴을 검출한 후 문자와 배경사이의 차이를 평가함으로써 후보 문자 요소들을 검출한다. 이후 후보 문자 요소들의 유사성을 계산하여 이를 분류한 후 완화법(relaxational operation)을 사용하여 문자 요소들을 병합시킨다. 이 논문에서는 위 알고리즘을 교통 표지판, 자동차 번호판, 간판, 그리고 화물열차 등에 포함되어 있는 다양한 크기의 문자들을 그레이 영상으로 변환하여 실험하였다. 이 방법은 기울어진 문자의 추출률이 낮은 단점이 있다. M. A. Smith[11] 는 비디오 프레임으로부터 문자 정보를 추출해내기 위한 기법을 제공한다. 이 논문에서는 문자들이 수평방향으로 놓여져 있다고 가정을 한다. 알고리즘은 먼저 높은 명암대비에 의해 특성화되어진 영역들을 추출하기 위하여 임계값을 적용하고 이후 수직 성분을 검색하는 단계를 거친다. 다음으로 비문자 요소들을 제거하고 첫 단계 처리에 의한 손상으로 인해 분리되어진 영역들을 통합하기 위하여 Smoothing 필터를 적용한다. 마지막으로 필터링 단계를 거치는데 후보 영역들이 본 알고리즘에서 정한 임계값

에 따라 충분히 큰 외곽 사각형을 가지고 있는 것, 외곽 사각형의 화소들이 임계값에 따라 충분히 밀집되어 있는 것, 그리고 장축 대 단축의 비율이 정한 임계값에 따라 수평방향이 수직방향 보다 큰 것 등에 따라 후보 문자 영역들이 선택되어 진다. 그리고 부분적인 밝기 히스토그램을 분석하여 낮은 명암 대비를 가지고 있는 영역들은 제거한다. 본 방법에서는 다른 명암 대비를 가지고 있는 문자들은 잘 추출되지 않는 단점이 있다. Zhong *et al.*[12] 은 컬러 영상에서 문자를 추출하기 위한 두 가지 방법과 다시 이 두 방법을 통합하여 사용하는 방법을 제안하고 있다. 첫 번째 방법은 컬러 영상에 컬러 양자화를 적용하고 연결 요소 분석을 각 양자화 컬러 면상으로 적용한다. 이후 필터링 과정을 통해서 비문자 요소들을 제거해 나간다. 두 번째 방법은 그레이 영상에서 공간적인 분산을 사용하는데 이는 배경의 분산 값이 문자 영역의 분산 값보다 낮다는 가정에 기반한 것이다. 이 방법은 올라가거나 내려간 문자는 잘 추출되지 않는다는 단점이 있다. H. K. Kim[13]은 컬러 비디오 프레임으로부터 자동으로 문자열을 추출하기 위한 방법을 제안하고 있다. 이 논문에서는 문자들이 수평방향으로 놓여져 있고 균일한 컬러와 일정한 크기를 가지고 있다고 가정한다. 알고리즘은 크게 두 부분으로 구성되어져 있다. 첫 번째 단계에서 컬러 히스토그램의 양자화에 의해 컬러 영상을 분할한다. 가장 지배적인 컬러를 색상 클러스터링에 의해 분할하고 결과 영역들의 수는 임계값에 의해 좌우되어 진다. 두 번째 단계는 휴리스틱 필터링의 적용이다. 첫 번째 휴리스틱 필터링은 영상의 가장자리에 닿아 있는 영역들과 수평 방향으로 너무 긴 요소를 포함하고 있는 영역들을 제거한다. 다음으로 두 번째 휴리스틱 필터링으로 이 논문에서는 X-시그네처와 Y-시그네처에 대한 분석을 하는데 이는 영상에서 수평방향(X)과 수직방향(Y)으로 각각 프로젝션 하여 후보 문자 요소들을 색상 면(plane)별로 적용한다. 이후 검출되어진 문자열의 영상과 시공간상의 위치 등이 데이터 베이스에 저장되어진다. 이 논문에서는 알고리즘의 결과가 제시하고 있는 16가지의 임계값에 의존적이다. 한 예로, 색상 면의 수가 아주 중요한 요소가 되는데 너무 많은 면은 문자열의 부분이 다른 면에 지정되어 질 수도 있고 너무 적은 면은 다른 요소가 하나로 통합되어 질 수도 있기 때문이다. P. K. Kim[14] 은 컬러 영상에서 문자열의 위치를 찾기 위한 방법을 제안하고 있다. 이 논문에서는 전체 영상에 대해 컬러 양자화를 적용하지 않고 지역적으로 적용한다. 알고리즘은 크게 네 부분으로 구성된다. 첫 번째는 입력영상을 256 색상의 영상으

로 변환한다. 두 번째는 지역적 컬러 양자화를 이용하여 경계선 추적(contour following)을 적용한다. 세 번째는 연결 요소를 추출해내고 마지막으로 필터링 과정을 거쳐 최종적으로 문자열의 위치를 찾는다. 이 논문에서는 지역적 컬러 양자화를 적용함으로써 전역적 컬러 양자화를 적용했을 때 발생할 수 있는 문자 영역과 배경 영역의 병합을 배제하여 문자열의 위치 검출율을 높여주는 하나 지역적 컬러 양자화를 모든 컬러에 적용함으로써 많은 계산 량을 요구하게 된다.

3. 제안된 방법

본 논문에서 제안된 방법은 그림 1에서 보여진 것처럼 선 처리, 후보 문자 영역 추출, 필터링 등의 세 단계로 구성된다. 선 처리 단계에서는 컬러 영상을 그레이 영상으로 변환한 후 영상의 대비를 강화시키기 위하여 히스토그램 스트레칭 방법을 적용하고 이후 변형된 지역적 이진화 방법을 적용한다. 후보 문자 영역 추출 단계에서는 변형된 지역적 이진화가 적용된 결과 영상에 모폴로지 연산을 적절하게 사용하여, 우선 문자 영역은 아니면서 문자로 판단되기 쉬운 양의 오류(false-positive) 요소들이 강조되어 남아있는 영상을 만든다. 또한, 변형된 이진화 결과 영상에 모폴로지 연산과 본 논문에서 제안한 기하학적 보정 필터링 방법을 적용하여 문자와 문자로 판단되기 쉬운 요소들이 모두 강조되어 남아있는 영

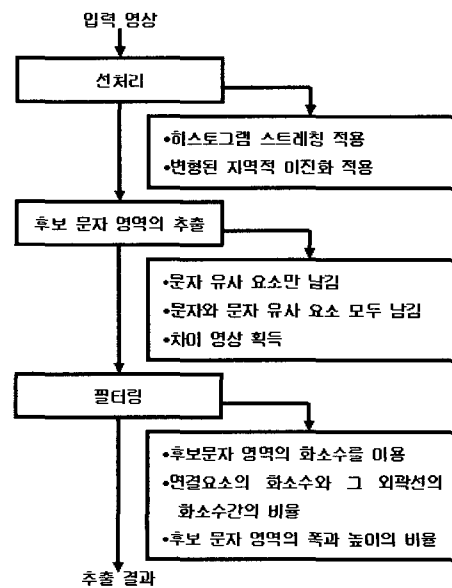


그림 1 제안된 방법의 절차

상을 만든다. 이 두 영상의 차를 구함으로써 찾고자 하는 문자 요소들이 주로 남고, 문자가 아닌 문자처럼 보이는 오류 요소들은 대부분 제거된 결과 영상을 만든다. 문자로 판단되는 양의 오류 영역들을 남기는데 사용된 모폴로지 연산은 3x3 크기의 구조 요소를 갖는 열림과 (열림닫힘 + 닫힘열림)/2 이며, 문자 및 문자와 유사한 요소들을 남기는데 사용된 연산은 (열림닫힘 + 닫힘열림)/2와 기하학적 보정이다. 마지막으로 필터링 단계에서는, 각 후보 문자 영역의 특성들을 이용하여 비문자 요소들 제거한다. 본 논문에서 추출하고자 하는 문자 영역은 최소한 연속된 세 개의 문자를 갖는 가로 영역으로 가정한다.

3.1 선 처리 단계

문자 추출 분야에서는 선 처리 단계에서 입력 영상을 컬러 영상 자체로 사용하거나 그레이 영상으로 변환하여 사용하는 두 가지의 방법이 있다. 문자 추출 분야에 있어서 일반적 논문의 경향에서 나타난 입력 영상의 사용 예를 보면 표 1에서처럼 Ohya et al.[9], Lee and Kankanhalli[10], Smith and Kanade[11], Lienhart and Stuber[15] 등의 논문에서는 입력 영상을 그레이 영상으로 변환하여 선 처리에 사용하며 Kim[13], P. K. Kim[14]의 논문에서는 입력 영상을 컬러 영상으로 선 처리에 사용한다. 또한 Zhong et al.[12]의 논문에서는 입력 영상을 컬러 영상과 그레이 영상으로 혼성하여 사용한다. 입력 영상을 컬러 영상으로 사용할 경우 대개 컬러 양자화를 적용하여 색의 연속성을 이용하는데 이 방법은 유사한 색상들이 선택되어진 지배적인 색상으로 합쳐 질 수 있기 때문에 문자 영역과 배경 영역이 병합될 수 있다는 단점이 있다 P. K. Kim[14]. 또한 문자 영역과 배경에서 밝기는 거의 같은 값을 가질 수 있는 반면 각 컬러 요소들 R, G, B는 다른 값을 가질 수 있다. 따라서 본 논문에서는 문자 추출을 위해 그레이 영상을 사용하고 문자와 배경사이의 밝기 대비와 문자의 형태를 사용한다. 입력 영상의 R, G, B값은 아래 식 (1)을 통해 그레이 영상으로 변환된다.

$$Y=0.299R+0.587G+0.114B \quad (1)$$

표 1 선 처리를 위한 영상의 유형과 특성

	그레이 영상을 사용하는 방법	컬러 영상을 사용하는 방법
종류	Ohya et al.[9] Lee and Kankanhalli[10] Smith and Kanade[11] Lienhart and Stuber[15]	Zhong et al.[12] Kim.[13] P. K. Kim[14]
특징	이진화, 문자영역과배경영역의 밝기 대비 이용	컬러 양자화, 색의 연속성 이용

다음으로 그레이 영상으로 변환된 입력 영상의 대비를 강화시키기 위하여 히스토그램 스트레칭 방법을 적용하는데 이를 통해 문자 영역과 배경 사이의 밝기 차이를 강조하게 된다. 밝기 대비가 강화된 영상을 (영상의 폭/30)*(영상의 높이/30)의 블록으로 분할한 후 변형된 지역적 이진화 방법을 적용한다. 적용적 임계값 T는 식 2에서 보여진 것처럼 각 블록 화소 값의 평균(m)과 표준편차(σ)에 의해서 계산된다. 사용자 입력 변수 k는 뉴스, 스포츠, 영화, 광고영상 등과 같은 입력영상의 종류에 따라 조절되어 질 수 있다. 본 논문에서 사용하는 변형된 지역적 이진화 방법은 지역적 이진화방법의 변형된 형태로서 영상의 화소 값이 임계값보다 작을 경우는 영으로 변환하고 임계값보다 크거나 같은 경우는 영상이 가지고 있는 그레이 값을 그대로 유지하는 방법이다. 이는 후보 문자 영역 추출단계에서 모폴로지 연산을 그레이 영상에서 적용하기 위함이다. 그림 6-(c)는 변형된 지역적 이진화 방법의 결과를 보여주고 있다.

$$T = m + \sigma * k \quad (2)$$

3.2 후보 문자 영역 추출

3.2.1 문자 유사 요소만 남김

본 논문에서 문자 요소란 영상에 포함되어 있는 정확히 문자를 포함하는 요소를 말하고, 문자 유사 요소란 문자 요소는 아니지만 문자 요소로 추정되기 쉬운 요소들을 말한다. 본 과정에서는 모폴로지 열림과 (열림닫힘+닫힘열림)/2 연산이 변형된 지역적 이진화결과 영상에 적용된다. 먼저 변형된 지역적 이진화 결과 영상의 화소 값이 영을 갖지 않는 화소들을 255로 변환함으로써 이진화 영상을 만들고 이 영상에 3x3크기의 구조 요소를 사용하여 열림 연산을 적용한다. 3x3크기의 구조 요소는 뉴스 비디오에 나타나는 문자들의 폭을 조사하여 선택한 것이다. 이 열림 연산에서는 녹입이 잠음과 문자 유사요소들을 제거하게 되고 블립이 녹입 연산과정에서 손상 입은 요소들을 복구시키게 된다. 다음으로 열림 연산이 적용된 결과인 이진 영상에서 255값을 가지는 화소에 변형된 지역적 이진화 결과 영상의 같은 픽셀 위치의 밝기 값들을 복사한다. 이 결과 영상에 그림 2와 같은 1x5 크기의 구조 요소를 사용하여 (열림닫힘+닫힘열림)/2 연산을 적용한다. 본 논문에서의 (열림닫힘+닫힘열림)/2 연산은 그레이 영상에 적용되는 모폴로지 연산으로서, 입력 영상에 열림과 닫힘을 차례로 적용한 영상과, 다시 입력 영상에 닫힘과 열림을 적용한 후, 이 두 영상의 평균을 취하는 것이다 (식 3 참조).

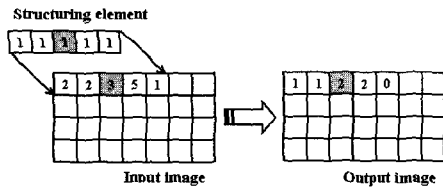
$$(OpenClose + CloseOpen)/2 = \frac{(I \cdot S) \cdot S + (I \cdot S) \cdot S}{2} \quad (3)$$

- I: 입력 영상
- S: 구조 요소
- : 열림 연산
- : 닫힘 연산

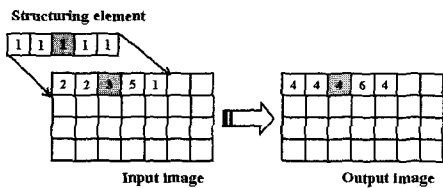


그림 2 (열림닫힘+닫힘열림)/2를 위한 구조 요소

본 논문에서 적용한 그레이 영상에 대한 (열림닫힘+닫힘열림)/2에 사용한 녹임 및 불림 방법은 아래 그림 3과 같다. 먼저 녹임의 경우 구조 요소를 입력 영상과 일치시킨 후, 구조 요소에 대응되는 영상의 화소 값의 차를 구한다. 예를 들면 입력 영상이 그림 3-(a)와 같이 주어지면 [2-1, 2-1, 3-1, 5-1, 1-1]의 결과인 [1, 1, 2, 4, 0]에 대하여 중심의 화소 값 2보다 작은 값들은 그대로 두고, 큰 값들은 중심 화소 값으로 대체하는 것이다. 불림의 경우는 녹임과 같은 과정을 거친 후, 구조 요소에 대응되는 영상의 화소 값의 합을 구한다. 예로서 그림 3-(b)에서처럼 [2+1, 2+1, 3+1, 5+1, 1+1]의 결과인 [3, 3, 4, 6, 2]에 대하여 중심의 화소 값 4보다 큰 값들은 그대로 두고, 작은 값은 중심 화소 값으로 대체한다.



(a) 그레이 영상에서의 녹임 연산

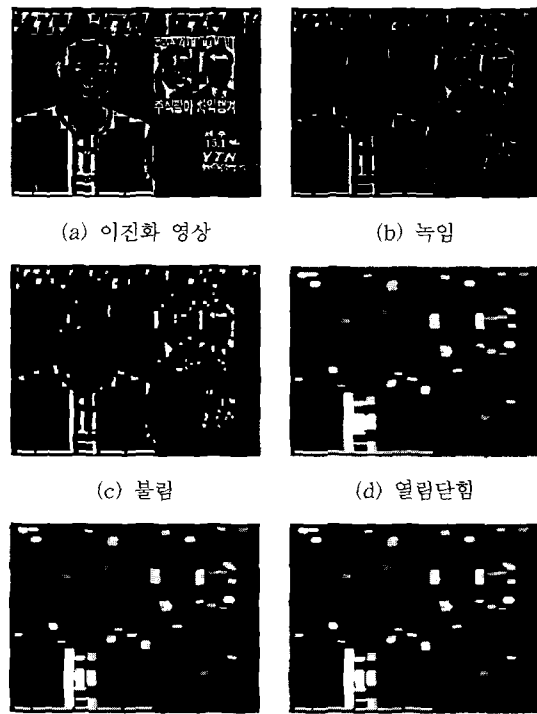


(b) 그레이 영상에서의 불림 연산

그림 3 본 논문에서 사용한 그레이 영상에서의 모폴로지 연산

그레이 영상에서의 모폴로지 연산 (열림닫힘+닫힘열림)/2는 열림닫힘에 의해 배경에 존재하는 잡음을 감소

시키고 닫힘열림에 의해 각 요소의 영역 내에 존재하는 빈 공간을 메워주는 효과가 있다 [18]. 이 (열림닫힘+닫힘열림)/2 연산은 1x5의 구조 요소를 사용하여 오직 수평 방향으로만 적용하는데 이는 뉴스 비디오에 나타나는 대부분의 문자들이 수평 방향으로 존재하기 때문이다. 그림 4는 본 단계에서의 각 모폴로지 연산의 결과를 보여주고 있다.



(a) 이진화 영상 (b) 녹임 (c) 불림 (d) 열림닫힘 (e) 닫힘열림 (f) (열림닫힘+닫힘열림)/2

그림 4 각 모폴로지 연산의 결과

3.2.2 문자와 문자 유사 요소 모두 남김

본 과정에서는 (열림닫힘+닫힘열림)/2 연산과 기하학적 보정방법이 적용된다. 먼저 배경에 존재하는 잡음을 감소시키고 각 요소의 영역 내에 존재하는 빈 공간을 메우기 위해 (열림닫힘+닫힘열림)/2 연산이 적용된다. 이후 이 결과 영상의 화소 값이 영을 갖지 않는 화소들을 255로 변환함으로써 이진화 영상을 만들어 기하학적 보정 필터링을 적용한다. 그림 5에서 보여진 것처럼 기하학적 보정은 모폴로지 연산에 의해 손상을 받은 후보 문자 요소들을 보존하고 보상하는데 좋은 효과를 준다. 이 연산은 수평과 수직 방향으로 모두 적용되는데 수평

방향의 임계값이 수직방향보다 크다. 이는 뉴스 영상에 나타나는 문자는 대부분 수평방향으로 존재하기 때문이다. 수평방향의 임계값은 20이 수직방향의 임계값은 5가 사용되었다. 본 논문에서 제안한 기하학적 보정은 화소들간의 간격을 임계값에 의해 연결해주거나 유지시켜주는 역할을 한다. 그림 5에서 해당 화소의 값이 255일 경우 연속된 화소를 따라 가다가 새롭게 0인 화소를 만나게 되면 다시 255의 화소를 만날 때까지 0인 화소수를 세게 된다. 그리고 중간에 존재하는 0의 연속적인 화소수가 임계값 이하이면 0을 255로 만들어 주어 두개의 단절된 선을 이어주고 임계값 이상이라면 원래 0의 값으로 유지시킨다.

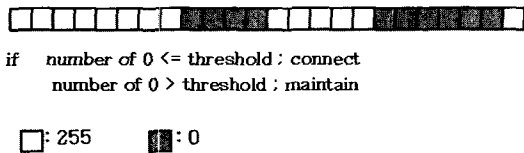


그림 5 기하학적 보정의 예

3.2.3 차이 영상 획득

모폴로지 연산을 사용해서 얻은 결과인 문자 유사 요소만 남김 단계의 결과인 이진 영상은 오직 문자 유사 요소들만을 포함하고 있다. 그리고 문자 유사요소와 문자 요소 모두 남김의 결과인 이진영상은 문자요소와 문자 유사 요소 모두를 포함하고 있다. 이 두 영상의 차를 구함으로써 찾고자 하는 문자 요소들이 주로 남고, 문자가 아닌 문자처럼 보이는 오류 요소들은 그림 6-(e)와 같이 대부분 제거된 결과 영상을 만든다.

3.3 후보 문자 영역의 필터링

본 과정에서는 비문자 요소와 잡음이 제거되어 진다. 먼저 연결 요소 방법이 적용되어지고 이후 필터링 과정이 적용된다. 본 논문의 필터링 과정에서 사용된 임계값들은 실험에 의해 가장 우수한 성능을 보여주는 값들을 선택한 것으로 비율과 같이 상대적인 값이므로 영상의 크기가 달라지더라도 같은 결과 값을 보여 줄 수 있다.

먼저 각 연결 요소에서의 후보 문자 영역의 화소수를 이용한 방법을 적용한다. 각 연결 요소(C)에서 전체 영상 픽셀 수 대비(NP_i) 4%이하의 화소수를 가진 후보

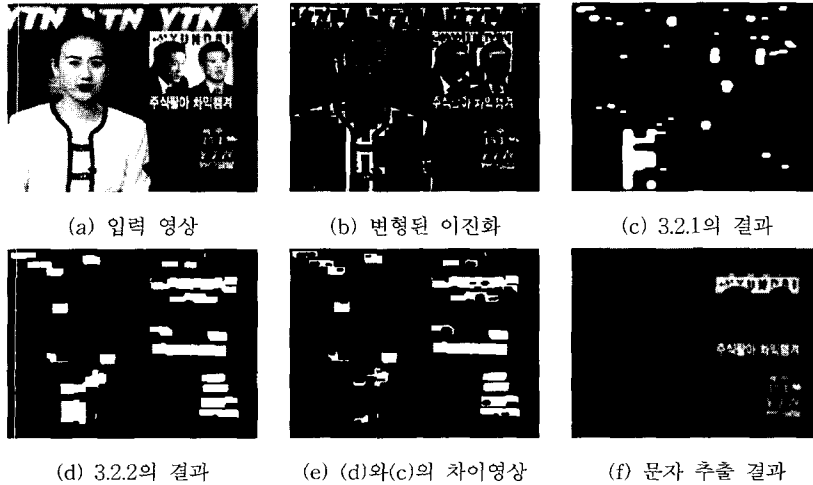


그림 6 선 처리와 문자 영역 추출과정의 결과

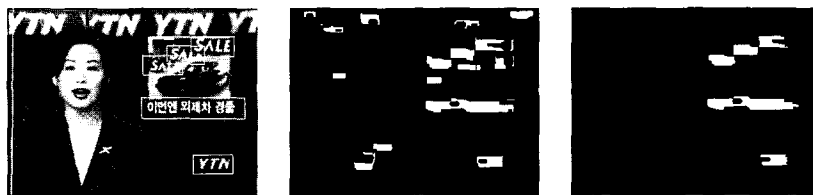


그림 7 후보 문자 영역의 화소수를 이용한 방법의 결과

문자 영역은 제거된다. 본 논문은 연속된 세 개 이상의 문자 영역을 최소의 문자 영역으로 가정하고 있기 때문에 뉴스에 나타나는 문자들의 평균 폭을 조사하여 전체 영상에서 차지하는 비율을 이용한 것으로 연속된 세 문자들의 평균 화소 비율은 전체 영상 화소의 4% 보다 작다는 실험적인 결과에 기반한 것이다. 이 방법을 적용하여 얻어진 결과가 그림 7에 보여진다.

$$\begin{aligned} \text{if } NP_i \leq 0.04 \text{ then } C_i = \text{제거} \\ \text{else } C_i = \text{유지} \end{aligned} \quad (4)$$

다음으로 각 연결 요소의 화소수와 그 외곽선의 화소수간의 비율을 이용한 필터링 방법을 적용한다. 각 연결 요소(C_i)에서 후보 문자 영역과 그 외곽선의 화소수 비율(NB_i)이 0.23이하인 요소들은 제거된다. 이는 문자 요소로 추정되는 연결 요소의 전체 화소수와 외곽선의 화소수는 일정한 비율 이하라는 실험에 기반한 것이다. 즉 문자로 추정되는 연결 요소와 비문자로 추정되는 연결 요소의 크기와 외곽선의 길이를 조사하여 얻은 값이다. 이 방법을 적용하여 얻어진 결과 영상이 그림 8에 보여진다.

$$\begin{aligned} \text{if } NB_i \leq 0.23 \text{ then } C_i = \text{제거} \\ \text{else } C_i = \text{유지} \end{aligned} \quad (5)$$

마지막으로 후보 문자 영역의 폭과 높이의 비율을 이용한 필터링 방법을 적용한다. 각 연결 요소(C_i)에서 후보 문자 영역에 외곽 사각형을 만들고 외곽 사각형의 폭($Width$)대 높이($Height$)의 비율이 0.66이하인 요소들은 제거된다. 이는 본 논문에서 한 문자라인은 최소한 세 문자 이상으로 이루어져 있다는 가정과 뉴스 영상에 나타나는 문자는 대부분 가로 방향으로 존재한다는 가

정에 기반한 것으로 뉴스 영상에 나타나는 문자들의 폭과 높이의 비율을 조사하여 얻은 결과이다. 수직 방향의 문자는 이 임계값을 조정함으로써 추출되어질 수 있다. 이 방법을 적용하여 얻어진 결과 영상이 그림 9에 보여진다.

$$\begin{aligned} \text{if } Height/Width \leq 0.66 \text{ then } C_i = \text{제거} \\ \text{else } C_i = \text{유지} \end{aligned} \quad (6)$$

4. 실험 결과 및 분석

본 논문에서 제안된 알고리즘은 866MHz 펜티엄 프로세서상의 개인용 컴퓨터에서 마이크로 소프트웨어++ 6.0을 사용하여 구현되어 졌다. 제안된 방법의 평가를 위해서 300개의 국내 뉴스 영상(MBC, YTN, KBS)과 100개의 국외 뉴스 영상(CNN, BBC, Bloomberg, Arirang)을 사용하였다. 또한, 서로 다른 종류의 비디오 상에서 문자 추출 결과와 비교하기 위하여 자막이 포함된 100개의 영화 비디오 영상과 50개의 TV 광고에 대해서도 실험하였다. 영상의 크기는 320*210이며 각 영상은 연속적인 프레임에서 임의로 선택되어졌다. 문자 추출 결과를 평가하기 위하여 식 7과 같이 세 가지 성능 기준을 정하여 사용하였다. 첫 번째는 정확 추출률로써 문자를 정확하게 추출해낸 경우이다. 두 번째는 실용 추출률로써 약간의 오차는 포함하고 있지만 자동 비디오 인덱싱에 사용될 수 있는 키워드를 포함하고 있는 경우이다. 마지막 세 번째는 오류율로써 문자 요소가 아님에도 문자로 인식한 경우와 존재하는 문자를 추출해내지 못한 경우 모두이다.



그림 8 연결 요소의 화소수와 그 외곽선 화소수간의 비율을 이용한 방법의 결과



그림 9 후보 문자 영역의 폭과 높이의 비율을 이용한 방법의 결과

정확 추출률 = N_{ct} / T_{nt} (7)

실용 추출률 = $(N_{ct} + N_{pt}) / T_{nt}$

오류율 = $(N_{nr} + N_{mt}) / T_{nt}$

T_{nt} : 실험 영상에서 존재하는 전체 문자의 수

N_{ct} : 정확하게 추출된 문자의 총수

N_{pt} : 허용 범위 내에서 추출된 문자의 총수

N_{nr} : 비 문자 영역을 문자로 추출한 총수

N_{mt} : 문자 추출에 실패한 총수

그림 10-(a)는 정확 추출의 결과로서 문자열을 모두 포함하여 찾은 경우이다. 그림 10-(b)는 실용 추출의 예로서 약간의 추출 오류를 포함하고 있지만, 찾은 문자열이 비디오 인덱싱에 사용될 수 있는 키워드를 포함하고, 문자열의 위치를 찾은 경우이다. 부분적인 추출 오류는 모폴로지 연산이 적용되는 과정에서 손상된 요소들이 제대로 복구되지 않았기 때문에 발생한다. 그림 10-(c)는 비 문자요소를 문자요소로 틀리게 추출한 경우로서 원인은 추출된 비문자 요소의 전경과 배경의 명암 대비가 크기 때문에 변경된 지역적 이진화 단계에서 제거되지 않았기 때문이다. 그림 10-(d)는 문자요소 추출에 실패한 경우로서 이는 문자와 배경 사이의 명암 대비가 낮아서 변경된 지역적 이진화 단계에서 이러한 요소가 제거되었기 때문이다. 제안한 방법에서 오류의 발생의 가장 큰 원인은 영상에서 문자가 너무 작거나 문자와 배경의 밝기 대비가 너무 낮은 경우, 그리고 동일 문자열에서 각 문자의 거리가 너무 멀리 떨어져 있는 경우 등이다. 표 2는 각 실험 영상에 대한 전체 결과를 나타낸 것이다.

국내 뉴스와 국외 뉴스를 분류하여 비교한 이유는 한글은 각 문자의 높이가 일정하지만 영어는 올라간 문자

표 2 실험 영상에 대한 문자 추출의 전체 결과

	정확 추출률	실용 추출률	오류율
국내 뉴스	533, 77.5%	(533+110), 93.6%	(46+42), 12.8%
국외 뉴스	291, 76.5%	(291+43), 88.1%	(6+45), 13.4%
영화 비디오	132, 73.7%	(132+43), 97.7%	(2+4), 3.3%
광고 영상	66, 49.3%	(66+29), 70.9%	(8+39), 35.1%

와 내려간 문자로 인해서 변형된 지역적 이진화 과정에서 오류가 발생할 수 있기 때문에 인식률의 차이가 있을 수 있기 때문이다. 위의 실험결과에서 볼 수 있듯이 국내 뉴스의 문자 추출률이 국외 뉴스의 문자 추출률보다 조금 높음을 알 수 있다. 이러한 문제는 입력 영상의 컬러 정보를 이용함으로써 개선되어질 수 있다. 영화 비디오에서는 문자의 크기와 대비가 뉴스 영상보다 일정한 점은 있으나 배경의 복잡도가 훨씬 높은 이유로 인해 전체적으로 실용 인식률은 뉴스 비디오 영상보다 좋은 결과를 얻을 수 있으나 정확 인식률에 있어서는 인식률이 낮아짐을 알 수 있었다. 광고 영상에서는 문자의 크기, 서체, 방향, 위치, 빛의 변화 정도, 그리고 배경의 복잡도 등이 일반 뉴스 영상보다 크기 때문에 결과가 좋지 않음을 알 수 있었다. 이는 광고 영상의 특성상 배경 문자가 일반 뉴스 영상보다 훨씬 많이 포함되어 있으며 또한 문자에 다양한 색상과 많은 빛의 변화 그리고 복잡한 배경으로 인해 변형된 지역적 이진화 단계에서 이를 효과적으로 분할해내지 못한 것에 기인한 것이다. 따라서 본 논문에서 제안한 방법은 뉴스 비디오와 같은 영상에 포함되어 있는 인위적으로 삽입된 문자들에 대해서는 높은 인식률을 얻을 수 있었으나 복잡한

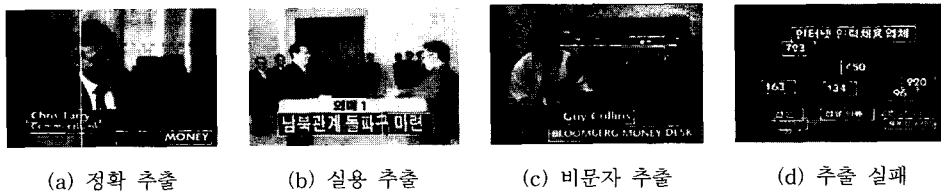


그림 10 각 문자 추출 결과 평가 기준의 예시 영상

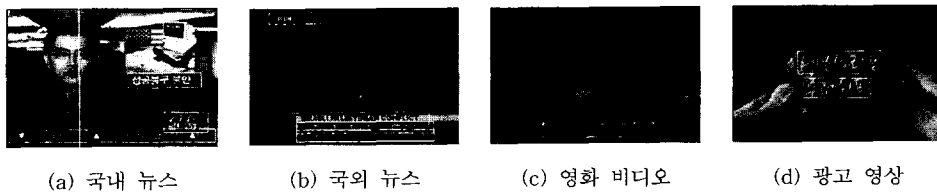


그림 11 각 실험영상에 대한 문자 추출 결과

배경을 가진 실의 영상에서 자연적으로 포함되어진 배경 문자에 대해서는 인식률이 낮음을 알수 있다.

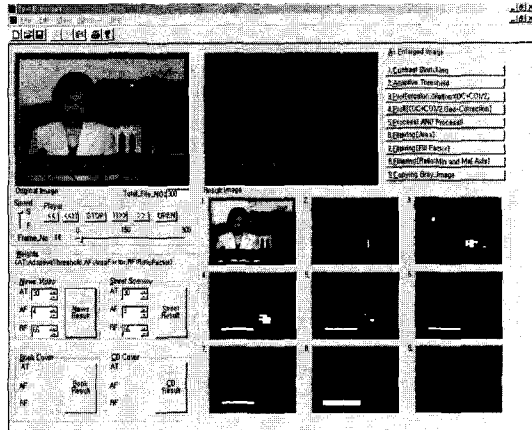


그림 12 본 논문에서 제안하는 문자 추출 시스템 인터페이스

5. 결론 및 향후 연구 과제

본 논문에서는 디지털 비디오의 영상으로부터 문자를 추출하기 위한 새로운 방법을 제안하였다. 제안한 방법은 적응적 이진화가 적용된 결과 영상에 모폴로지 연산을 적절하게 사용하여, 우선 문자 영역은 아니면서 문자로 판단되기 쉬운 양의 오류 요소들이 강조되어 남아있는 영상을 만들고 또한, 변형된 이진화 결과 영상에 모폴로지 연산과 본 논문에서 제안한 기하학적 보정 필터링 방법을 적용하여 문자와 문자로 판단되기 쉬운 요소들이 모두 강조되어 남아있는 영상을 만들어 이 두 영상의 차를 구함으로써 찾고자 하는 문자 요소들이 주로 남고, 문자가 아닌 문자처럼 보이는 오류 요소들은 대부분 제거된 결과 영상을 획득함으로써 문자를 추출하는 방법이다. 문자로 판단되는 양의 오류 영역들을 남기는데 사용된 모폴로지 연산은 3x3 크기의 구조 요소를 갖는 열림과 (열림닫힘+닫힘열림)/2 이며, 문자 및 문자와 유사한 요소들을 남기는데 사용된 연산은 (열림닫힘+닫힘열림)/2와 기하학적 보정이다. 열림닫힘과 닫힘열림 연산은 잡음을 감소시키고 영역 내에서 존재하는 빈 공간을 채우는데 효과적이다. 또한 본 논문에서 제안한 기하학적 보정 필터링은 모폴로지 연산에 의해 손상을 받은 후보 문자 요소들을 보존하고 보상하는데 좋은 효과를 준다. 본 논문의 실험결과는 제안한 문자 추출 알고리즘이 디지털 뉴스 비디오 영상에 존재하는 문자의 추출에 우수한 성능을 나타내고 있음을 보여준다. 또한 다

른 다양한 영상에서도 문자의 폭에 따라 모폴로지 연산의 구조 요소를 조정함에 의해서 좋은 결과를 얻을 수 있다. 앞으로의 연구는 본 알고리즘을 책표지, 컴팩트 디스크 표지, 그리고 거리의 풍경 영상등에서도 유연한 결과를 보여 줄 수 있도록 지속적으로 연구 할 것이다.

참고 문헌

- [1] Jae-Chang Shim, Chitra Dorai, Ruud Bolle, "Automatic Text Extraction from Video for Content-Based Annotation and Retrieval," Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on, On page(s): 618 - 620 vol.1 16-20 Aug. 1998
- [2] Anil K. Jain, Bin Yu, "Automatic text location in images and video frames," Pattern Recognition, Vol. 31, No. 12, pp. 2055-2076, 1998.
- [3] H.Kuwano, Y.Taniguchi, H.Arai, M.Mori, S.Kuraka-ke, H.Kojima, "Telop-on-demand: video structuring and retrieval based on text recognition," Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on, On page(s): 759 - 762 vol.2 30 July-2 Aug. 2000
- [4] U. Gargi, S. Antani, R. Kasturi, "Indexing text events in digital video databases," Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on, On page(s): 916 - 918 vol.1 16-20 Aug. 1998
- [5] Sameer Antani, Ullas Gargi, David Crandall, Tarak Gandhi and Rangachar Kasturi, "Extraction of Text in Video," Dept. of Comput. Sci. & Eng., Pennsylvania State Univ., Technical Report, CSE-99-016, August 30, 1999
- [6] S. Messelodi and C.M. Modena, "Automatic identification and skew estimation of text lines in real scene images," Pattern Recognition, Vol. 32 (5) (1999) pp. 791-810
- [7] S. Antani, D. Crandall, R. Kasturi, "Robust extraction of text in video," Pattern Recognition, 2000. Proceedings. 15th International Conference on, Volume: 1, 2000, Page(s): 831 -834 vol.1
- [8] Y. Lu, "Machine printed character segmentation- An overview," Pattern Recognition 28, 1995, 67-80
- [9] J. Ohya, A. Shio, S. Akamatsu, "Recognizing characters in scene images," IEEE Trans. Pattern Anal. Mach. Intell. PAMI-16(2) (1994) 214-220.
- [10] C. M. Lee, A. Kankanhalli, "Automatic extraction of characters in complex images," Int. J. Pattern Recognition Artificial Intell. 9 (1) (1995) 67-82.
- [11] M. A. Smith, T. Kanade, "Video skimming for quick browsing base on audio and image

- characterization," Technical Report CMU-CS-95-186, Carnegie Mellon University. July 1995.
- [12] Y. Zhong, K. Karu, A. K. Jain, "Locating text in complex color images," *Pattern Recognition*, 28 (10) (1995) 1523-1535.
- [13] H. K. Kim, "Efficient automatic text location method and content-based indexing and structuring of video database," *J. Visual Commun. Image Representation* 7 (4) (1996) 336-344.
- [14] Pyeoung-Kee Kim, "Automatic Text Location in Complex Color Images using Local Color Quantization", *TENCON 99. Proceedings of the IEEE Region 10 Conference*, Volume: 1, pp. 629-632, 1999.
- [15] R. Lienhart, F. Stuber, "Automatic text recognition in digital videos," *Image and Video Processing IV 1996, SPIE 2666-20*, 1996.
- [16] M. Bertini, C. Colombo, A. Del Isimbo, "Automatic Caption Localization in Videos using Salient Points," *IEEE Int. Conf. on Multimedia and Expo.* pp69-72, 2001
- [17] F. LeBourgeois, "Robust Multifont OCR System from Gray Level Images," in *International Conference on Document Analysis and Recognition*, vol. 1, pp.1-5, 1997.
- [18] J. Serra, *Image Analysis and Mathematical Morphology*. New York: Academic, 1982.



김길천

2001년 경기대학교 전자계산학과 졸업(이학사). 2001년 ~ 현재 연세대학교 대학원 컴퓨터학과 석사과정 재학중. 관심분야는 영상검색, 영상처리, 비디오 처리, 패턴 인식



변혜란

1980년 연세대학교 수학과 졸업(이학사). 1983년 연세대학교 대학원 수학과 졸업(이학석사). 1987년 Univ. of Illinois, Computer Science(M.S.). 1993년 Purdue Univ., Computer Science(Ph.D.). 1994년 ~ 1995년 한림대학교 정보공학과 조교수. 1995년 ~ 1998년 연세대학교 컴퓨터과학과 조교수. 1998년 ~ 현재 연세대학교 컴퓨터과학과 부교수. 관심분야는 인공지능, 영상인식, 영상처리



장인영

2000년 경기대학교 전자계산학과 졸업(이학사). 2000년 ~ 현재 연세대학교 대학원 컴퓨터학과 석사과정 재학중. 관심분야는 영상검색, 비디오 인덱싱, 패턴 인식, 인공지능



고병철

1998년 경기대학교 전자계산학과 졸업(이학사). 1998년 ~ 2000년 연세대학교 대학원 컴퓨터학과 졸업(공학석사). 2000년 ~ 현재 연세대학교 대학원 컴퓨터학과 박사과정 재학중. 관심분야는 영상검색, 비디오 인덱싱, 패턴 인식, 인공지능

공지능