

텐서보팅을 이용한 텍스트 배열정보의 획득과 이를 이용한 텍스트 검출

(Extraction of Text Alignment by Tensor Voting and its Application to Text Detection)

이 귀 상 [†] 또 안 ^{**} 박 종 현 ^{***}
(GueeSang Lee) (Toan Nguyen Dinh) (JongHyun Park)

요약 본 논문에서는 이차원 텐서보팅과 에지 기반 방법을 이용하여 자연영상에서 문자를 검출하는 새로운 방법을 제시한다. 텍스트의 문자들은 보통 연속적인 완전한 곡선 상에 배열되어 있고 서로 가깝게 위치하며, 이러한 특성은 텐서보팅에 의하여 효과적으로 검출될 수 있다. 이차원 텐서보팅은 토큰의 연속성을 curve saliency 로 산출하며 이러한 특성은 다양한 영상해석에 사용된다. 먼저 에지 검출을 이용하여 영상 내의 텍스트 영역이 위치할 가능성이 있는 텍스트 후보영역을 찾고 이러한 후보영역의 연속성을 텐서보팅에 의해 검증하여 잡음영역을 제거하고 텍스트 영역만을 구분한다. 실험 결과, 제안된 방법은 복잡한 자연영상에서 효과적으로 텍스트 영역을 검출함을 확인하였다.

키워드 : 텐서보팅, 텍스트 검출, 자연영상, 영역기반 텍스트 검출, 커브 돌출도

Abstract A novel algorithm using 2D tensor voting and edge-based approach is proposed for text detection in natural scene images. The tensor voting is used based on the fact that characters in a text line are usually close together on a smooth curve and therefore the tokens corresponding to centers of these characters have high curve saliency values. First, a suitable edge-based method is used to find all possible text regions. Since the false positive rate of text detection result generated from the edge-based method is high, 2D tensor voting is applied to remove false positives and find only text regions. The experimental results show that our method successfully detects text regions in many complex natural scene images.

Key words : tensor voting, text detection, natural scene image, region-based text detection, curve saliency

· 이 논문은 2008년도 정부(교육과학기술부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구(KRF-2008-313-D00999)이며 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음(NIPA-2009-(C1090-0903-0008))

· [†] **통신회원** : 전남대학교 전자컴퓨터공학부 교수
gslee@jnu.ac.kr

^{**} **학생회원** : 전남대학교 전자컴퓨터공학부
toan_mulmi@hotmail.com

^{***} **정회원** : 전남대학교 전자컴퓨터공학부 연구교수
jhpark@jnu.ac.kr
(Corresponding author)

논문접수 : 2009년 7월 10일

심사완료 : 2009년 9월 19일

Copyright©2009 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 소프트웨어 및 응용 제36권 제11호(2009.11)

1. 서론

최근 디지털 카메라나 휴대폰 카메라를 통해 막대한 양의 영상자료들이 생성되고 있다. 자연 영상에서 문자를 이해하는 것은 시각 장애인용 보조 장치 또는 외국인을 위한 통역 시스템같이 매우 중요하고 다양한 용도로 활용될 수 있다. 텍스트 영역 검출은 텍스트 인식을 위한 첫 번째 단계이다. 자연 영상은 다양한 조건에서 촬영될 수 있기 때문에 원하는 텍스트영역만을 검출하는 것은 도전적인 연구 영역이다. 자연 영상에서 배경은 매우 복잡하게 마련이며 그 속에 존재하는 텍스트 역시 다양한 폰트, 크기, 방향을 갖고 있다[1]. 이뿐 아니라 일정치 않은 조명, 반사, 그림자, 그리고 하이이라이트 같은 경우는 텍스트 인식을 더 어렵게 한다[2].

본 논문에서의 텍스트 검출은 영상 내에서 텍스트를 감지하고 텍스트의 위치를 찾는 것을 말한다. 최근 자연

영상에서 텍스트를 검출하는 다양한 알고리즘이 제시되었다. 보통 텍스트 검출을 위해서는 자연 영상의 저차원 정보를 사용한다. 여기서 저차원 정보란 색 연속성, 그레이 레벨 분산, 색 분산 정보를 말한다. 특히 색 연속성은 대부분의 텍스트 영역 내 텍스트 색이 동일한 특징을 갖기 때문에 매우 유용하게 쓰인다. 또한 텍스트가 갖는 고유한 성질 때문에 그레이 레벨 분산과 색분산 정보를 통해 텍스트를 배경과 분리할 수 있다. 텍스트 검출에 대한 응용 및 관련 시스템에 대해서는 많은 연구가 이루어지고 있다[3,4].

텍스트 검출 방법들은 크게 텍스처 기반 방법과 영역 기반 방법의 두 가지로 나눌 수 있다. 텍스처 기반 방법은 미분 등의 연산이나 주파수 기반 접근에 의해 높은 계산량을 요구할 수 있다. [1]에서는 지역적인 특징성분을 제안하고 신경망 분류기를 사용하여 스케일이나 3D 방향에 대해 불변특성을 갖는 영상 내 텍스트 영역의 위치 검출방법을 제안하였다. 주파수 변환을 이용한 방법은 푸리에 변환[5], 이산 코사인 변환[6], 웨이블릿[7]을 사용한다. 주파수 도메인을 이용한 방법들은 작은 텍스트들이 텍스처 특징에 대해 강한 응답성을 갖는다는 특성에 기반하고 있다. 이러한 이유 때문에 이런 방법들은 작은 문자들을 검출에 있어서 더 나은 성능을 보인다. 그러나 텍스트 문자의 폰트의 크기나 형태에 따라 이러한 주파수 특성이 달라질 수 있으며 이러한 변화에 적용하기 어려운 단점을 갖고 있다.

영역기반 방법은 에지 정보나 문자 획의 연결요소 특징 같은 공간적 응집성을 이용한다[2, 8, 9]. 에지 기반 방법에서는 텍스트 외곽선으로부터 에지를 검출하고 이를 병합한 후 몇 가지 휴리스틱 요소를 사용하여 비텍스트 영역을 걸러낸다. 이와는 반대로 연결요소 기반 방법에서는 이진화를 통해 발생한 각 연결요소를 이미지 내의 모든 영역에 존재하는 텍스트가 검출될 때까지 그룹화해 나가는 bottom up 접근이다. 요소 병합을 위해서 이완 알고리즘이 사용되기도 한다[10]. 그러나 연결요소 기반의 검출 결과는 잡음영역과 텍스트의 구분이 쉽지 않고, 텍스트의 경사와 기울어짐에 의해 영향을 받는다. [11]에서는 디지털 이미지에서 자동으로 문자 영역을 추출하는 방법이 제시되었으며, 이는 단일 색상의 텍스트에 대해 분할 및 병합(split and merge) 알고리즘을 사용하였다. 영역 기반 접근법에서는 주변 정보를 고려하는 것이 쉽지 않기 때문에 많은 비텍스트 영역이 텍스트 영역으로 검출되는 오류가 발생하기 쉽다.

이전에 제안된 방법 중에 특히 영역 기반 접근에서는 자연 영상 내에 글자와 비슷한 특징을 가진 부분들이 많은 경우 오인식률이 높았다. 이러한 오인식률을 줄이고 글자들의 크기와 방향이 일정치 않은 조건에서도 강

건한 검출을 보장하기 위해서 본 논문에서는 조명 변화에 강건한 새로운 영역 기반 방법으로 2D 텐서보팅과 에지 기반 접근 방법을 이용한 텍스트 검출 알고리즘을 제안한다[12,13].

텐서보팅은 입력 토큰을 대칭텐서로 표현하고 토큰들 간의 통신을 위한 보팅 알고리즘으로 이루어진, 컴퓨터 비전 분야의 다양한 문제를 다룰 수 있는 통합된 추론 체계이다. 보통 영상 내의 텍스트는 이를 구성하는 문자들이 좁은 간격으로 일정한 형태로 배열되는 특징이 있다. 따라서 텍스트의 에지검출을 활용하여 문자영역 후보를 검색한 후, 2D 텐서보팅을 적용하면 글자의 위치를 찾거나 노이즈의 제거, 그리고 오검출을 방지하기 위한 중요한 정보를 생성할 수 있다. 지금까지 텐서보팅을 텍스트 처리에 적용한 예는 [14,15]에서 찾아볼 수 있다. 이러한 연구결과에서는 주로 텐서보팅을 활용하여 텍스트 영상의 잡음을 제거하고 왜곡을 보정하여 텍스트를 분할하는 내용이며, 텍스트 검출과는 직접적인 관계를 갖는 것은 아니다.

본 논문은 다음과 같은 순서로 구성되어 있다. 2장에서는 2D 텐서보팅에 대해서 간략한 소개와 기본적인 연산에 대한 설명을 하고, 3장에서는 제안한 텍스트 영역에 대한 검출에 대해 설명한다. 4장에서는 실험 결과에 대해 설명하며 마지막으로 5장에서 결론과 향후 연구에 대해 기술한다.

2. 2D 텐서보팅

텐서보팅은 USC(University of Southern California)의 컴퓨터 비전 그룹에 의해 개발된 통합 프레임워크이다[12,13]. 텐서보팅은 2D, 3D, 심지어 N차원의 희미하고 잡음이 섞인 이진 데이터에서 표면이나 경계, 또는 곡선이나 곡선의 교차점에 대한 표시를 생성하는데 사용된다.

텐서는 벡터의 개념을 확장한 기하학적인 양으로 균일한 원(circle) 또는 구체(sphere)에 한 방향의 어떠한 힘을 작용시키면 변형하여 타원 또는 타원체의 형태로 표현할 수 있다. 따라서 텐서를 구성하는 각각의 파라미터들은 입력 데이터의 특성을 나타내며 정해진 필드 내에서 보팅을 통하여 인접하는 정보와 교류 및 통신을 통하여 정교한 텐서로 재표현할 수 있다. 일반적으로 텐서보팅은 두 성분에 기반하고 있다. 하나는 데이터 표현을 위한 텐서 계산, 그리고 데이터들과 통신을 위한 보팅이다. 입력 데이터는 각각의 위치에서 이웃하는 데이터들에 자신의 정보를 전파(propagation)한다. 그 정보는 텐서로 인코딩하며 보팅필드에 의해 패턴 구조가 결정되어 진다. 따라서 각각의 위치에서 수집되어진 정보는 보팅필드내의 모든 보트들에 캐스트(cast)하여 각각

의 특징 형태를 위하여 텐서의 고유값(eigenvalue)을 이용한 saliency 맵을 만들 수 있다. 그리고 관심 대상이 되는 특징에 대하여 분석 및 추론을 수행할 수 있다[12,16].

2.1 2차 대칭 텐서의 표현

입력 데이터들은 토큰으로 표현할 수 있으며 각각의 토큰들은 2D 텐서로 계산 가능하다. 따라서 텐서로부터 기하학적인 특징 추출에 의한 패턴을 추론할 수 있으며 텐서들은 이웃하는 정보와 통신을 통하여 새로운 텐서로 재표현 할 수 있어 두 포인트 사이의 가장 적합한 경로를 만들 수 있다. 모든 토큰들은 이웃하는 토큰들로부터 보트(vote)들을 축적하며 수집된 정보는 2D 대칭 텐서의 합에 의해 다음 식 (1)과 같이 공분산 행렬로써 표현할 수 있다.

$$cov = \begin{pmatrix} \Sigma v_x v_x & \Sigma v_x v_y \\ \Sigma v_y v_x & \Sigma v_y v_y \end{pmatrix} \quad (1)$$

여기서 $[v_x, v_y]$ 는 주어진 데이터를 위하여 이웃하는 데이터에 의해 발생된 스틱 보트의 벡터이다. 주어진 고유 시스템은 두 개의 고유값 (λ_1, λ_2) 와, 두 개의 고유벡터는 (\hat{e}_1, \hat{e}_2) 로 구성되어 있으며 행렬 cov는 다음 (2) 식과 같이 표현할 수 있다.

$$cov = (\hat{e}_1 \hat{e}_2) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \hat{e}_1 \\ \hat{e}_2 \end{pmatrix} \quad (2)$$

$$cov_{2d} = (\lambda_1 - \lambda_2) \hat{e}_1 \hat{e}_1^T + \lambda_2 (\hat{e}_1 \hat{e}_1^T + \hat{e}_2 \hat{e}_2^T) \quad (3)$$

여기서, $\hat{e}_1 \hat{e}_1^T$ 와 $\hat{e}_1 \hat{e}_1^T + \hat{e}_2 \hat{e}_2^T$ 는 2D에서 스틱(stick) 텐서와 볼(ball) 텐서를 나타내며, 각각의 saliency는 $\lambda_1 - \lambda_2$ 와 λ_2 이다.

2.2 보팅필드

보팅필드는 2D 스틱 보팅필드로부터 유도할 수 있으며 일반적으로 이것을 2D 스틱 커널이라 부른다. 보팅필드에서 보트들은 접선 벡터나 법선 벡터에 의해 하나의 방향 성분으로 정의할 수 있다. 보트 필드내에서 보트들은 각각의 위치에서 접선 벡터 또는 법선 벡터에 의존한다. 그림 1에서, “공간에 주어진 O와 P를 지나는 곡선에 가장 부드러운 경로의 법선 벡터는 무엇인가?”에 대하여 생각해 볼 수 있다.

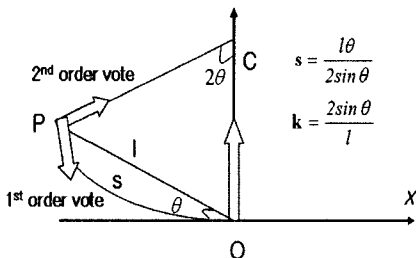


그림 1 스틱 텐서에 의한 2차 보팅필드

여기서 가상의 연속하는 호(arc)를 따라 일정한 곡선을 유지하기 때문에 접촉 서클에 연속하는 O와 P는 가장 적당한 것 중에 하나이다. 가장 적당한 법선은 점 P에서 곡면에 있는 법선에 의해 주어진다. 따라서 스틱 보트의 방향은 O와 P를 가장 부드럽게 연결하는 법선이 된다. P에서 법선 벡터의 길이는 보트의 saliency를 나타내는 호의 길이 OP와 circular arc의 곡면에 반비례한다. 따라서 모든 보팅 크기는 기본적인 2D 커널에 기반하고 있으며 다음 식 (4)과 같이 나타낼 수 있다.

$$SDF(s, k, \sigma) = e^{\left(\frac{s^2 + ck^2}{\sigma^2} \right)} \quad (4)$$

s는 OP의 호 길이이며, k는 곡률, c는 높은 곡률을 가지고 제어하기 위한 상수 값이다. 그리고 sigma는 효율적인 이웃 영역에 대한 크기로 분석을 위한 스케일 값을 나타낸다. 이 함수에 대한 자세한 내용은 [12]에 나타나 있다. 각각의 입력 위치에서 정규 방향의 벡터를 얻을 수 있다면, 우리는 2D 공간에서 높은 saliency 값을 가지고 있는 토큰들을 연결하기 위한 부드러운 구조를 추론할 수 있다. 특징 추출은 모든 불연속인 위치에서 보팅에 의해 수행된다.

2.3 텍스트 검출을 위한 2D 텐서보팅의 예

본 절에서는 간단한 예를 통해 2D 텐서보팅의 적용을 설명하고자 한다. 먼저 입력토큰으로서 그림 2와 같이 4개의 픽셀 P1(5, 5), P2(10, 18), P3(10, 22), P4(10, 25)가 20x40 영상에 존재한다고 하자. 단, 영상의 기준점은 왼쪽 하단으로 한다.

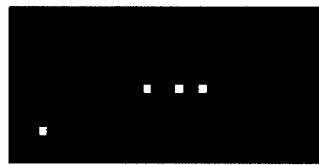


그림 2 입력토큰의 예

각각의 토큰은 볼 tensor로서 표 %와 같이 2x2 행렬로 나타내어질 수 있다. 이는 2.1절에서 자세히 설명하였다. 표 1에서 e1, e2는 각각 normal vector와 tangent vector를 나타내며, lambda1, lambda2는 텐서표현의 eigenvalue를 나타낸다. 초기값에서는 curve saliency 값이 모두 0으로 설정된다. 또한 e1, e2는 임의의 값으로 설정되어도 관계없으며, 텐서보팅 결과 최종 벡터값을 얻게 된다.

보팅필드 매개변수 sigma = 8으로 가정하고 볼 보팅필드는 다음과 같이 나타내어질 수 있다.

이제 그림 4와 같이 각 입력 토큰을 각각의 볼 보팅필드로 대치하여 보팅을 실시한다. 보팅필드에 기반하여

표 1 텐서보팅을 위한 초기값 설정

입력토큰	P1(5,5)	P2(10,18)	P3(10,22)	P4(10,25)
e_1	$(0,1)^t$	$(0,1)^t$	$(0,1)^t$	$(0,1)^t$
e_2	$(0,1)^t$	$(0,1)^t$	$(0,1)^t$	$(0,1)^t$
λ_1	1	1	1	1
λ_2	1	1	1	1
Curve saliency ($\lambda_1 - \lambda_2$)	0	0	0	0
행렬	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

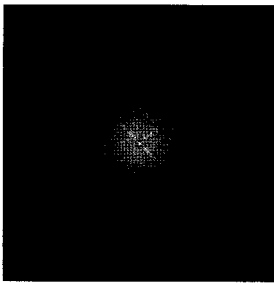


그림 3 볼 텐서의 보팅필드

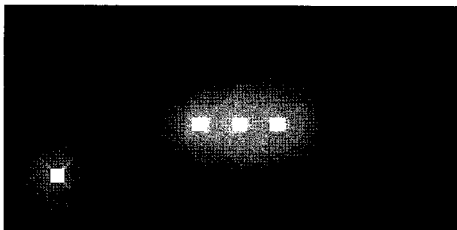


그림 4 입력토큰에 보팅필드를 곱친 영상

보팅된 값은 원래의 행렬에 가산되어 누적되며 최종 행렬이 일반 텐서를 나타내며 이제는 다른 normal vector (e_1)와 curve saliency를 갖게 된다.

표 2의 결과에서 입력토큰 P1은 고려된 점이므로 텐

표 2 텐서보팅 후의 텐서 정보 변화

입력토큰	P1(5,5)	P2(10,18)	P3(10,22)	P4(10,25)
e_1	$(0,1)^t$	$(1,0)^t$	$(1,0)^t$	$(1,0)^t$
e_2	$(0,1)^t$	$(0,1)^t$	$(0,1)^t$	$(0,1)^t$
λ_1	1	1.74	1.82	1.75
λ_2	1	1.3	1.2	1.3
Curve saliency ($\lambda_1 - \lambda_2$)	0	0.44	0.62	0.45
행렬	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1.5 & 0.0 \\ 0.1 & 0.7 \end{pmatrix}$	$\begin{pmatrix} 1.5 & 0.1 \\ 0.1 & 1.8 \end{pmatrix}$	$\begin{pmatrix} 1.5 & -0.0003 \\ -0.0003 & 1.7 \end{pmatrix}$

서보팅결과 아무런 값의 변화를 갖지 못하며 curve saliency 값도 0으로 남아 있다. 반면 나머지 연결된 점들 P2, P3, P4는 서로 보팅필드에 의한 보팅값들이 누적되어 행렬이 변화되며 이러한 행렬로부터 위의 벡터와 eigenvalue를 구할 수 있다. 이 점들은 curve saliency가 상당히 높게 나타남을 볼 수 있다. 이러한 형태의 텐서보팅은 본 논문의 텍스트 후보 영역의 문자 중심점에 대해 수행하여 일련의 직선형태의 배열인지를 판단하게 된다.

3. 제안한 텍스트 검출 알고리즘

3.1 제안한 알고리즘 개요

그림 5에 제시된 텍스트 검출 알고리즘은 크게 텍스트 후보영역 추출과 텐서보팅에 의해 추정된 텍스트 문자 중심점의 배열 추출, 그리고 이러한 두 가지 정보를 통합하여 최종적으로 텍스트 영역을 검출하는 단계로 구성되어 있다. 일반적으로 에지검출에 의한 텍스트 후보영역 추출은 잡음영역과의 구분이 쉽지 않으며, 본 논문에서는 이러한 문제점을 극복하기 위하여 텐서보팅에 의한 텍스트 배열로 추정되는 문자 중심점을 찾아내어 이를 텍스트 후보영역과 결합하여 최종결과를 생성한다.

텍스트 배열추정을 위해서 텐서보팅 입력 토큰 생성과 2D 텐서보팅의 적용에 의한 curve saliency 계산이 필요하며, 이를 위해 먼저 입력이미지로부터 변환된 그레이 영상에 에지 검출 방법을 적용하여 연결요소를 생성한다. 이 연결요소는 다시 텍스트의 특징을 이용하여 텍스트로 구분될 가능성이 있는 텍스트 후보영역으로 구분된다. 본 논문에서는 텍스트가 비교적 수평형태로 직선 또는 완만한 곡선상에 배열된다고 가정하며 이러한 텍스트 배열추정을 위해 수직성분 검출 Sobel 연산자를 사용하여 텐서보팅을 위한 입력 토큰을 생성한다.

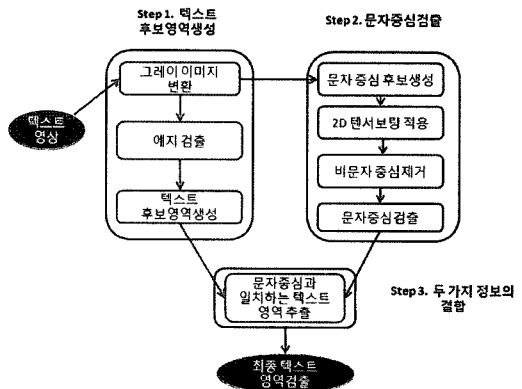


그림 5 텍스트 검출 알고리즘 순서도

3.2 텍스트 후보 영역 추출

먼저 텍스트 후보 영역을 생성하기 위해 에지 검출 방법을 사용한다. 주어진 칼라영상은 그레이 영상으로 변환된 후 Canny 에지 검출 알고리즘을 적용하여 이진 에지 영상으로 변환된다[17]. 그림 6(c)은 에지 검출 결과를 보여주고 있으며, 여기에는 많은 비 텍스트 영역이 포함되어 있음을 알 수 있다. 그림 6(d)은 연결요소 중에서 텍스트 영역으로 구분되기 어려운 크기를 가진 에지성분을 제거한 결과이다.

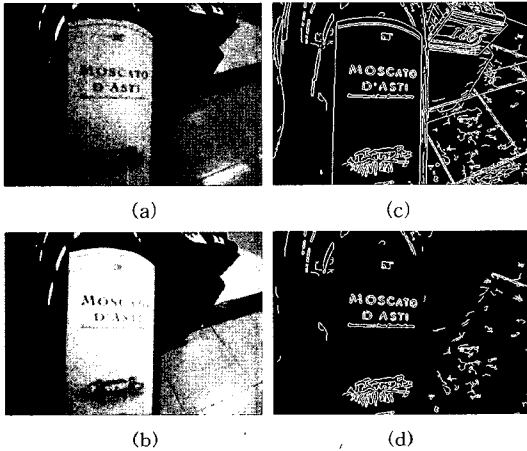


그림 6 연결요소 분석을 통한 텍스트 후보영역: (a) 원 영상, (b) 그레이영상, (c) Canny 에지 검출 영상, (d) 과도한 크기의 연결요소를 제거한 결과

3.3 텐서보팅을 위한 입력 정보생성

텐서보팅을 위한 입력으로서 텍스트 문자의 중심점으로 추정되는 점을 생성한다. 여기서는 각 문자들의 중심점을 찾아야 하므로 앞 단계에서 사용한 Canny 에지검출방법은 적절치 않다. 본 논문에서는 문자들이 대체로 수평성분을 갖는 직선상에 배열되어 있거나 수평방향의 완만한 곡선 형태로 배열되어 있다고 가정한다. 따라서, 각 문자가 다수의 수직성분으로 구분될 수 있도록 수직 성분 검출 Sobel 연산을 사용한다. $f(x,y)$ 를 픽셀 (x,y) 의 명도라고 할 때, 수평 또는 수직성분은 다음 식으로 나타낼 수 있다.

$$\frac{\partial f}{\partial x} = f(x+1,y) - f(x,y), \quad \frac{\partial f}{\partial y} = f(x,y+1) - f(x,y) \quad (5)$$

수직성분은 다음의 Sobel 에지 검출 필터를 사용하여 검출할 수 있다.

$$\begin{vmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{vmatrix}$$

$$G_x = \frac{\partial f(x,y)}{\partial x} = -f(x-1,y-1) - 2 \times f(x-1,y) - f(x-1,y+1) + f(x+1,y-1) + 2 \times f(x+1,y) + f(x+1,y+1) \quad (6)$$

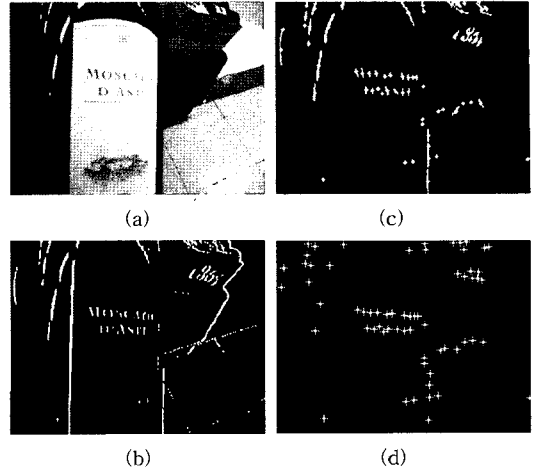


그림 7 (a) 그레이 영상, (b) 향상된 이진 에지 영상, (c) 비 텍스트 영역의 연결요소를 제거 후의 중심점 생성, (d) 최종 중심점 생성결과

그림 7(b)은 이러한 에지검출 결과를 나타내고 있으며 그림 7(c)은 크기가 적절치 않은 에지성분을 제거한 후의 연결요소 중심점을 보인 것이다. 최종적인 텐서보팅 입력으로 사용한 입력 토큰으로서의 중심점들은 그림 7(d)에 나타나 있다. 그리고, 남은 연결요소들의 중심점의 리스트를 생성한다. 그림 7(d)에서 확인할 수 있듯이 텍스트 영역의 중심점은 서로 가까이 위치하고 있으며 직선 또는 완만한 곡선 상에 위치하고 있다.

3.4 텍스트 영역의 문자배열 추정을 위한 텐서보팅

2장에서 언급했듯이 2D 텐서보팅 방법에서는 위에서 생성한 각각의 입력토큰은 텐서로 표현되게 된다. 텐서는 그들의 정보를 보팅 알고리즘에 의해 그 이웃들로 전파시킨다. 텐서보팅 후에 텍스트 영역 내의 중심점들에 대응하는 텐서들의 커브 saliency 들은 다른 토큰들에 비해서 높은 값을 갖는다. 그리고 또한 텍스트 영역의 토큰이 갖는 곡면 법선 벡터들은 텍스트 라인의 법선 벡터(normal vector)를 생성한다. 그림 8은 연결 요소들의 중심점과 함께 이 값을 시각화 하고 있다. 이 그림에서 중심점은 노란색 점으로, saliency 값은 흰색 선의 길이로, 토큰의 법선 벡터는 흰색의 방향으로 나타내었다. 그림에서 볼 수 있듯이 텍스트 영역 내에서는 토큰이 높은 saliency 값을 가지고 있고 또한 그들의 법선 벡터가 텍스트 라인의 법선 벡터와 비슷함을 알 수 있다. 그러므로 우리는 이 정보를 비 텍스트 영역을 제거

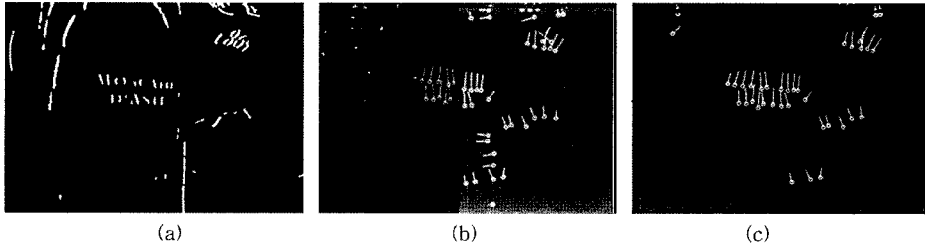


그림 8 (a) 연결요소 (b) 텐서보팅결과로 얻어진 토른 정보 (c) 비 텍스트 영역을 제거

하는데 사용할 수 있다. 작은 커브 saliency 값을 갖는 중심점이나 법선 벡터가 거의 수평 방향인 것은 제거된다. 이 단계는 그림 8에 나타내었다. 그림 8(c)는 토른 정보를 사용하여 다른 중심점을 제거한 뒤의 남은 중심점을 보여주고 있다. 남은 중심점들은 이전 단계에서 생성된 결과와 함께 최종 텍스트 검출을 위해 사용된다.

3.5 최종 텍스트 영역 검출

첫 번째 단계에서 생성된 후보 텍스트 영역과 두 번째 단계에서 생성된 중심점들을 활용하여 최종 텍스트 영역을 검출한다. 에지 기반 방법만으로는 많은 비 텍스트 영역이 텍스트 후보 영역으로 생성된다. 이러한 비 텍스트 영역을 제거하기 위해 중심점 정보를 활용한다. 중심점 정보를 사용하기 위해서 수평 방향에 가까운 모든 작은 크기의 연결요소들을 팽창 연산으로 연결한다. 그림 6(d)의 에지정보들은 팽창연산을 이용하여 텍스트 후보영역에 해당하는 연결요소 그룹으로 변환된다. 이러한 후보영역 그룹에 대해 그림 8(c)에서 구해진 중심점을 중첩한 것이 그림 9(a)이다. 이제 후보 영역에서 텍스트 영역으로 볼 수 없는 것을 제거한다. 먼저, 중첩된 중심점이 없거나 중심점의 수가 작아 문자 중심점의 연결로 간주될 수 없는 후보영역 또는 연결 요소는 제거

한다. 또한 연결요소가 너무 커서 도저히 텍스트 영역으로 간주할 수 없는 경우도 제거하게 된다. 그림 9(b)는 이러한 후보 영역 중에서 적절하게 배치된 중심점과 중첩되는 영역만을 텍스트 영역으로 선택한 결과이다. 최종 텍스트 검출 결과는 그림 9(c)에 나타내었다.

4. 실험 결과

제안된 방법의 성능을 평가하기 위해서 12개의 라벨 영상 및 2003년 ICDAR 테스트 이미지를 사용하였다.

모든 실험 영상의 해상도는 640 × 480이다. 또한 제시된 방법과 에지 기반 방법과 연결 요소 기반 방법과 비교하였으며, 이 두 가지 기존방법에 대한 코드는 [18]에서 제시된 코드를 사용하였다. Precision과 Recall은 다음 식을 이용하여 계산하였다.

$$Precision = \frac{T}{T+F} \times 100\% \tag{7}$$

$$Recall = \frac{T}{S} \times 100\% \tag{8}$$

여기서 S는 영상 내 텍스트 문자수, T는 검출에 성공한 텍스트 영역의 수, F는 오검출 영역의 수이다. 표 3은 제시한 방법이 다른 방법과 비교했을 때 많은 자연

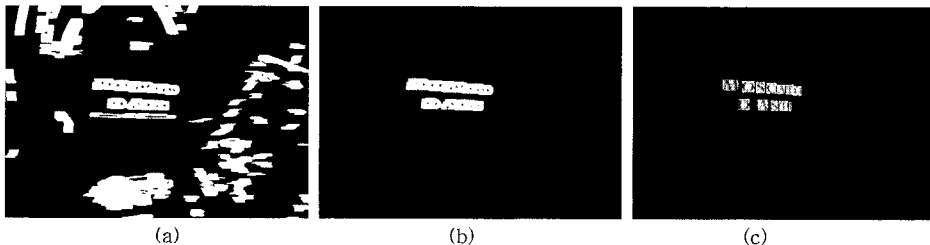


그림 9 (a) 팽창 적용된 연결요소와 중심점들, (b) 적은 수의 중심점을 가진 연결 요소 제거 (c) 최종 텍스트검출 결과

표 3 텍스트영역 검출 결과

	Proposed method	Edge-based method[2,6]	CC-based method[3]
S	1161	1161	1161
T	1016	908	880
F	77	391	450
Precision	92.96%	69.9%	66.17%
Recall	87.51%	66.17%	75.80%

표 4 제안한 알고리즘을 적용한 텍스트 검출 결과

Original images	Results	Original images	Results

영상에서 텍스트 영역을 성공적으로 검출하고 있음을 보여주고 있다. 검출한 결과 영상의 일부에 대해서는 아래 표 4에 표시하고 있다.

5. 결론

전통적인 연결 요소 방법은 문자후보영역의 검출에 대한 오류 비율이 높기 때문에 본 논문에서는 2D 텐서 보팅을 이용하여 자연 영상에서의 텍스트 검출 성능을 향상시켰다. 실험결과는 제시한 방법이 텍스트를 복잡한

자연 영상 이미지에서 다양한 폰트와 크기 방향에 강건한 검출임을 보여주고 있다. 추후 연구로서는 효율적인 잡음 감쇄를 위해 1차 보팅을 이용하는 방법과 왜곡된 텍스트 라인을 정렬하기 위해 곡면법선 벡터를 이용하여 텍스트 라인의 곡률을 추정하는 방법이 필요하다고 보여진다.

참고 문헌

[1] P. Clark and M. Mirmehdi, "Recognizing text in real scenes," *International Journal on Document Analysis and Recognition* 4, no.4, pp.243-257, 2008.

[2] G. Julinda, E. Ralph and F. Bernd, "A Robust algorithm for Text detection in images," *Proceedings of the 3rd international symposium on Image and Signal Processing and Analysis*, 2003.

[3] K. Jung, K.I. Kim, A.K. Jain, "Text information extraction in images and video: a survey," *Pattern Recognition*, pp.977-997, 2004.

[4] J. Liang, D. Doermann, and H. Li, "Camera-based analysis of text and documents: a survey," *International Journal on Document Analysis and Recognition* 7, pp.84-104, 2005.

[5] B. K. Sin, S. K. Kim, and B. J. Cho, "Locating characters in scene images using frequency features," *International conference on pattern recognition*, pp.489-492, 2002.

[6] D. Crandall, S. Antani, and R. Kasturi, "Extraction of special effects caption text events from digital video," *International Journal of Document Analysis and Recognition* 5, no.2-3, pp.138-157, 2005.

[7] Q. Ye, Q. Huang, W. Gao, and D. Zhao, "Fast and Robust Text Detection in Images and Video Frames," *Image and Vision Computing* 23, no.6, pp.565-576, 2005.

[8] J. Samarabandu and X. P. Liu, "An edge-based text region extraction algorithm for indoor mobile robot navigation," *International Journal of Signal Processing*, pp.273-280, 2006.

[9] L. Xiaoqing, S. Jagath, "Multiscale Edge-Based Text Extraction from Complex Images," *ICME*, pp.1721-1724, 2006.

[10] J. Hoya, A. Shio and S. Akamatsu, "Recognizing Characters in Scene Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.16, no.2, pp.67-82, 1995.

[11] R. Lienhart, F. Stuber, "Automatic Text Recognition in Digital Videos," *Image and Video Proc. IV, SPIE*, 1996.

[12] G. Medioni, M.S. Lee, C.K. Tang, "A Computational Framework for Segmentation and Grouping," *Elsevier, Amsterdam*, 2000.

[13] W.S. Tong, C.K. Tang, and G. Medioni, "First Order Tensor Voting, and Application to 3-D

Scale Analysis," *Proc. CVPR*, pp.175-182, 2001.

[14] Jaeguyn Lim, Jonghyun Park, Gérard G. Medioni, "Text segmentation in color images using tensor voting," *Image and Vision Computing*, vol.25, Issue 5, pp.671-685, 2007.

[15] J.H. Park, J.M. Yoo and G.S. Lee, "A Tensor Voting for Corrupted Region Inference and Text Image Segmentation," *International Multimedia Modelling Conference, LNCS*, vol.4351, pp. 751-761, 2007.

[16] J. Jia, C.K. Tang, "Inference of Segmented Color and Texture Description by Tensor Voting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.26, no.6, pp.771-786, 2004.

[17] J. Canny, "A Computational Approach To Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp.679-714, 1986.

[18] Sneha Sharma, "Extraction of Text Regions in Natural Images," *Master Thesis from Rochester Institute of Technology*, 2007.



이 귀 상

1980년 서울대학교 전기공학과 학사. 1982년 서울대학교 전기계산기공학과 석사. 1991년 Pennsylvania 주립대학 전산학 박사. 1984년~현재 전남대학교 전자컴퓨터공학과 교수. 관심분야는 멀티미디어통신, 영상처리 및 컴퓨터비전, 임베디드 시스템



Nguyen Dinh Toan

2005년 Received the B.S. degree in Information Technology from Vietnam National University Ho Chi Minh City, University of Natural Science. 현재 M.S. degree in computer science at Chonnam National University in Korea. 관심분야는 Video coding, Ubiquitous Computing



박 중 현

1994년 목포대학교 전자공학과 학사. 1997년 목포대학교 전자공학과 석사. 2001년 목포대학교 전자공학과 박사. 2002년~2004년 전북대학교 Post-Doc. 2004년~2006년 University of Southern California, USA, Post-Doc. 2006년~현재 전남대학교 BK21 연구교수. 관심분야는 영상처리, 패턴인식, 컴퓨터비전, 의료영상처리