다해상도 웨이블릿 변환과 써포트 벡터 머신을 이용한 자연영상에서의 문자 영역 검증

배 경 숙[†] · 최 영 우[††]

요 약

이미지에서 문자 추출은 영상을 이해하기 위한 가장 기초적이고 중요한 문제이다. 본 논문에서는 문자의 획 특징을 이용하는 통계적인 방법으로 문자 영역을 검증하는 방법을 제안한다. 제안하는 방법은 16×16 크기의 텍스트와 비텍스트 이미지를 웨이블릿(wavelet) 변환하여 문자의 획과 방향성을 표현하는 36차원의 특징을 추출한다. 추출된 특징 중 변별력이 높은 특징만을 선택하여 SVM(Support Vector Machine) 분류기를 구성한다. 분류기를 이용하여 16×16 크기의 윈도우로 검증 영역을 스캔하면서, 각각의 윈도우를 텍스트와 비텍스트로 분류하고 최종적으로 검증 영역의 텍스트 여부를 결정한다. 제안한 방법을 적용함으로써 텍스트와 유사하여 구별하기 어려운 비텍스트 영역을 검증할 수 있었다.

Text Region Verification in Natural Scene Images using Multi-resolution Wavelet Transform and Support Vector Machine

Kyungsook Bae[†] · Yeongwoo Choi[††]

ABSTRACT

Extraction of texts from images is a fundamental and important problem to understand the images. This paper suggests a text region verification method by statistical means of stroke features of the characters. The method extracts 36 dimensional features from 16×16 sized text and non-text images using wavelet transform - these 36 dimensional features express stroke and direction of characters - and select 12 sub-features out of 36 dimensional features which yield adequate separation between classes. After selecting the features, SVM trains the selected features. For the verification of the text region, each 16×16 image block is scanned and classified as text or non-text. Then, the text region is finally decided as text region or non-text region. The proposed method is able to verify text regions which can hardly be distin guished.

키워드 : 웨이블릿 변환(Wavelet Transform), SVM, 베이지안 에러율(Bayesian Error Rate), 검증(Verification)

1. 서 론

이미지나 동영상에 인위적으로 삽입하거나 자연적으로 포함한 텍스트들은 이미지의 내용을 함축적이고 구체적으로 표현하는 중요한 정보들이다. 이러한 정보들을 실시간에 찾아내어 인식한다면 시각 장애인을 위한 보행 안내, 로봇 자동주행 등에 활용할 수 있다. 문서 영상에서의 문자 추출은 많은 연구가 수행되어 이미 상용화된 제품들이 있는 반면, 자연영상에 포함되어 있는 문자들은 해상도가 낮고 글자 형태와 크기, 색상 등이 다양하기 때문에 이를 추출하고 인식하는 일은 어려운 문제이다. 최근에는 다양한 종류의 자동화 시스템 개발로 복잡한 배경을 갖는 자연영상에서의 문자 영역 추출 연구가 수행되고 있다. 문자 영역 추출 연구는 텍스트의 밝기 변화, 색 변화, 색 연속성 등의 특징을 이용한 연구와 문자의 통계적인 특징을 이용하는 연구로 구분하여 생각할 수 있다.

Jain 등은[1] 이진 이미지, 웹 이미지, 색 이미지 및 비디오 프레임 등의 네 종류의 이미지에서 텍스트를 추출하는 방법을 제안하였다. 이진 및 웹 이미지에 대해서는 텍스트의 밝기 값이 균일하다는 특징을 이용하였고, 색 이미지 및 비디오 프레임에 대해서는 색 연속성 특징을 이용하였다. 밝기 값이 균일하다는 특징을 이용하여 다중 값 분해(multi-valued decomposition)를 통해 전경과 배경을 분리한 후 연결요소를 분석하여 영역을 추출하였으며, 색 연속성을 이용하여 색 줄임을 수행한 후 역시 다중 값 분해를 통해 전경과 배경을 분리한 연결요소를 분석하여 영역을 추출하였다. 이 방법은 네 가지 종류에 대해 서로 다른 특징과 임계 값을 적용하기 때문에 종류별로 수동적인 실험을 수행하였다. 실험결과 이진 이미지, 웹 이미지, 비디오 프레임에서는 높은 추출률을 보인 반면, 색 이미지에 대해서는 저조한 추출률을 보였다. 또한, 대부분의 실험 이미지가 비디오 프레임

※ 본 연구는 숙명여자대학교 2002년도 교내연구비 지원에 의해 수행되었음.
† 정 회 원 : 한국전자통신연구원 지능형로봇연구단 연구원
†† 정 회 원 : 숙명여자대학교 정보과학부 교수
논문접수 : 2004년 1월 20일, 심사완료 : 2004년 6월 19일

에 국한되어 있는 단점이 있다.

Zhong 등은[2] 텍스트의 색은 일정하며, 명도이미지에서 텍스트 영역은 공간적 분산(spatial variance) 값이 크다는 특징을 이용하고, 두 방법을 순차적으로 결합한 방법을 제안하였다. 색을 이용한 방법에서는 색 양자화를 수행하여 색의 개수를 줄이고 각 색 면에 대한 연결요소를 분석하여 영역을 추출하였다. 명도이미지를 이용한 방법에서는 공간적 분산을 적용한 후 에지를 추출하여 서로 반대 방향을 가지는 에지 쌍을 찾음으로써 텍스트 영역을 추출하였다. 끝으로 두 방법을 순차적으로 결합하여 영역을 확정하였다. 다양한 종류의 스캔된 이미지로 제안한 방법을 실험한 결과 길이가 짧거나 색이 일정하지 않은 문자열, 수직 방향 또는 필기된 문자열 등에서 오류가 발생하는 단점이 있었다.

H. K. Kim은[3] 비디오 프레임으로부터 자동으로 텍스트 영역을 추출하기 위해서 문자들이 수평 방향으로 놓여져 있고, 균일한 색과 일정한 크기를 갖는다는 가정으로 색 연속성 특징을 이용한 방법을 제안하였다. 알고리즘은 크게 두 부분으로서 우선 색 히스토그램 양자화에 의해 색 이미지를 분할하고, 연결요소의 길이와 X 방향 및 Y 방향의 서명(signature)을 이용하여 각 색 면에서 비텍스트 요소들을 제거하였다. 50개의 비디오 프레임으로 제안된 알고리즘을 실험한 결과 86%의 추출률을 보였으나, 색의 대비가 크지 않은 텍스트와 크기가 작은 텍스트에 대해서는 잘 찾지 못하며, 16개의 경험적 임계 값을 정해야 하는 단점이 있다.

텍스트의 통계적 특징을 이용한 추출 연구로서 Li 등은[4] 웨이블릿 변환을 이용하여 텍스트의 획 특징을 추출하고, 신경망(neural network)을 이용한 학습으로 비디오 프레임에서의 텍스트 패턴을 분류하는 방법을 제안하였다. 다양한 크기의 텍스트를 추출하기 위해서 원이미지로부터 피라미드 이미지를 생성하여 다양한 해상도에서 텍스트를 추출하였다. 이 방법에서 사용된 신경망은 추정하고자 하는 목적함수를 효과적으로 반영하지 못해, 일반화 성능에서 문제점을 드러내고 있다.

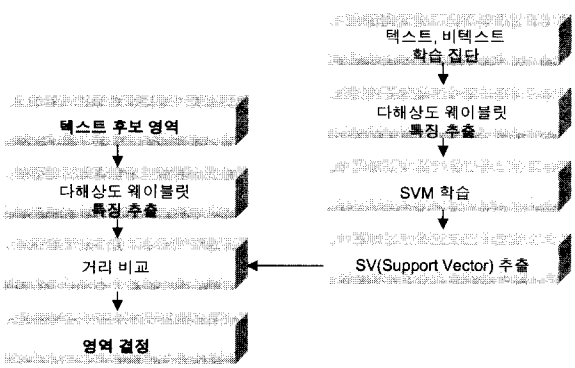
Kim은[5] SVM과 CAMSHIFT(Continuously Adaptive Mean Shift)를 이용한 텍스트 추출 방법을 제안하였다. 이 방법은 화소 자체의 밝기 값을 SVM에 입력하여 학습하므로 특징 추출에 소요되는 시간이 없으며 일반화 성능이 뛰어난 SVM을 학습 패러다임으로 사용하여 좋은 일반화 성능을 보였다. 그러나 SVM에 입력된 특징의 차원이 큰 것과 텍스트를 대표하는 특징이 정교하지 못한 단점이 있다.

Jeong 등은[6] 신경망을 이용하여 뉴스 비디오 프레임에서 텍스트의 위치를 찾는 방법을 제안하였다. 이 방법은 9×9 크기의 윈도우를 사용하여 80개의 이웃 화소와 1개의 중앙 화소를 신경망에 입력하여 중앙 화소의 텍스트 여부를 판별하고, 윈도우를 이동하면서 입력 이미지의 전체 화소를 텍스트 화소와 비텍스트 화소로 분류한다. 분류된 결과 이미지의 수평 및 수직 방향으로 텍스트 화소의 히스토그램을 구한 뒤 이를 분석하여 텍스트의 위치를 찾는다. 제안한 방법은 영어 및 숫자가 수평방향으로 정렬된 뉴스 비디오 프레임을 대상으로 하였기 때문에 텍스트의 크기와 폰트에 제한을 두었다.

본 논문에서는 웨이블릿 변환과[7] SVM을[8, 9] 이용하여 자연영상에서의 문자 영역 검증 방법을 제안한다. 제안한 방법은 웨이블릿 변환을 수행하여 36차원의 텍스트 획 특징을 추출하고, 그 중 변별력이 높은 12개의 특징만을 사용한다. 또한, 일반화 성능이 뛰어난 SVM을 이용하여 SV(Support Vector)들을 추출하였다. 검증 과정에서는 입력벡터와 SV와의 비교가 반복적으로 일어나므로 특징벡터의 차원이 낮을수록 검증 속도는 향상된다. 따라서 제안한 방법을 이용하여 비교적 일반화 성능이 뛰어나고 빠른 속도로 문자 영역을 검증할 수 있었다.

2. 제안 방법

문자 분류 또는 검증 문제는 일반적으로 특징 추출과 분류(classification) 문제로 나뉜다. 따라서 성공적인 문자 검증을 위해서는 텍스트와 비텍스트를 특징화할 수 있는 최적의 특징을 추출하고, 추출된 텍스트의 특징과 비텍스트의 특징을 식별할 효과적인 분류 패러다임을 세우는 것이 필요하다. 본 연구에서는 다해상도 웨이블릿(multi-resolution wavelet)을 이용하여 텍스트의 획 특징을 추출하고 추출된 특징을 SVM을 이용하여 학습하여 분류기를 만든다. (그림 1)은 본 논문에서 제안하는 문자 검증 시스템의 전체적인 개략도이다. 제안하는 방법은 특징 추출, SVM 학습, 검증의 세 단계를 거친다. 단계 1에서는 16×16 크기의 텍스트 영상과 비텍스트 영상 집단을 구성하고 다해상도 웨이블릿을 적용하여 텍스트의 획 특징을 추출한다. 단계2에서는 추출된 텍스트 영상과 비텍스트 영상의 특징을 SVM으로 학습시켜 분류기를 만든다. 단계 3에서는 색 정보나 명도 정보를 이용하여 가(假) 결정된 텍스트 후보영역을 16×16 윈도우로 스캔하면서 분류기를 이용하여 윈도우가 지나가는 위치의 블록을 텍스트와 비텍스트로 분류하고 최종적으로 영역에서 텍스트 블록이 차지하는 비율을 계산하여 영역의 텍스트 여부를 결정한다.



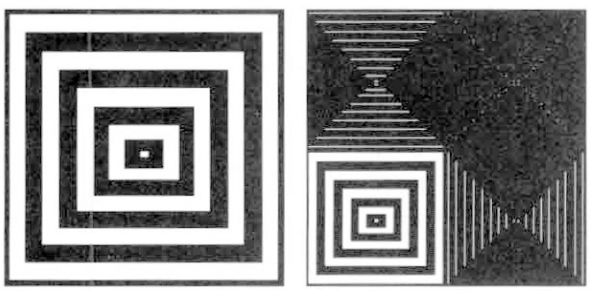
(그림 1) 제안하는 방법

3. 다해상도 웨이블릿 특징

본 장에서는 다해상도 웨이블릿 변환을 이용한 특징 추출에 대해서 설명한다. 3.1항에서는 웨이블릿 변환을 소개하고, 3.2항에서는 다해상도 웨이블릿 변환을 이용한 특징 추출 방법을 설명한다. 3.3항에서는 베이지안(Bayesian) 에러율을[10] 측정하여 추출된 특징 벡터를 낮은 차원의 벡터로 바꾸어 주는 방법을 제시한다.

3.1 웨이블릿 변환

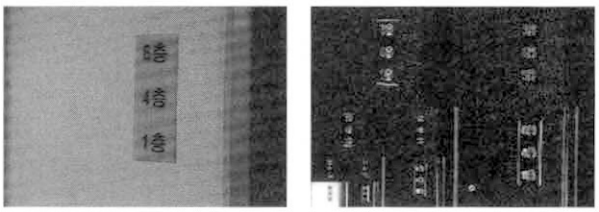
웨이블릿은 1983년 Morlet에 의해 소개된 이후 신호를 분석하고 해석하는데 효과적인 수학적 도구로 알려져, 순수 수학분야에서부터 여러 응용분야에 이르기까지 폭 넓게 연구되어 왔다. 웨이블릿 변환은[7] 영상에 수평, 수직 두 방향으로 필터를 적용하여 영상을 4개의 부밴드(subband) LH, HL, HH, LL로 분해한다. LH는 수평 방향으로 저주파 대역통과(lowpass) 필터를 적용한 뒤, 수직 방향으로 고주파 대역통과(highpass) 필터를 적용하여 영상의 수평 성분을 나타낸다. HL은 수평 방향으로 고주파 대역통과 필터를 적용한 뒤, 수직 방향으로 저주파 대역통과 필터를 적용하여 영상의 수직성분을 나타낸다. HH는 수평, 수직 두 방향으로 각각 고주파 대역통과 필터를 적용하여 영상의 대각 성분을 나타내며, LL은 수평, 수직 두 방향으로 각각 저주파 대역통과 필터를 적용하여 원 영상과 같은 통계적 특성을 나타낸다. 영상에 웨이블릿 변환을 수행한 뒤, LL 부밴드에 웨이블릿 변환을 반복적으로 적용하는 것을 "다해상도 웨이블릿 분해"라고 한다. (그림 2)는 영상에 웨이블릿 변환을 한 번 적용한 결과를 보여준다.



(그림 2) 원이미지에(좌) 대한 웨이블릿 변환 결과(우)

3.2 특징 추출

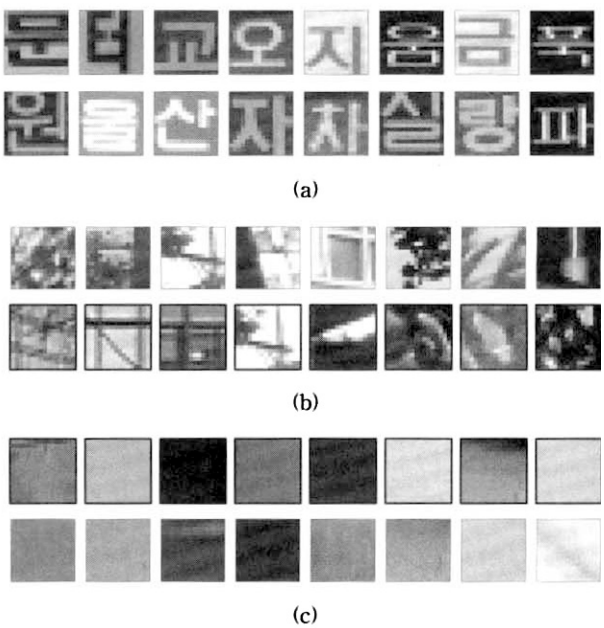
본 논문에서는 다해상도 웨이블릿을 이용하여 영상을 분해하여 특징을 추출한다. 웨이블릿은 여러 해상도로 영상을 분석하고, 고주파 대역 통과 필터링을 통해 고주파 성분인 에지를 검출한다. 또한, 저주파 대역 통과 필터링을 통과한 축소된 영상도 여전히 많은 양의 정보를 갖고 있기 때문에 정보의 양을 축소하는 데 유용하다. 웨이블릿 변환으로 생겨나는 4개의 부밴드 중 LH는 수평 고주파 성분, HL은 수직 고주파 성분, HH는 대각 고주파 성분에서 큰 웨이블릿 계수를 갖는다. (그림 3)은 영상에 3-레벨 웨이블릿을 적용하였을 때, 텍스트 영역이 세 개의 고주파 부밴드에서 두드러지게 나타남을 보여준다. 이러한 웨이블릿 변환의 특성은 영상에 웨이블릿 변환을 적용하였을 때, 텍스트의 특징인 수평, 수직, 대각선 방향의 획들이 고주파 부밴드에서 두드러지게 나타나도록 한다. 따라서 웨이블릿 변환을 이용하면 텍스트의 획을 표현하는 특징을 추출할 수 있다.



(그림 3) 원이미지에(좌) 대한 3-레벨 웨이블릿 변환 결과(우)

본 논문에서는 16×16 화소 크기의 영상에 웨이블릿 변환을 반복적으로 적용하여 특징 벡터를 추출한다. 이 때, 특징 벡터는 웨이블릿 변환으로 생겨난 부밴드들의 평균(M), 2차(μ_2), 3차(μ_3) *central moment*를 사용한다. 크기가 $N\times N$인 부밴드 I에 대해 특징들은 식 (1)과 같이 계산된다[4].

$$M(I)=\frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}I(i,j)$$
$$\mu_2(I)=\frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(I(i,j)-M(I))^2 \qquad (1)$$
$$\mu_3(I)=\frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(I(i,j)-M(I))^3$$



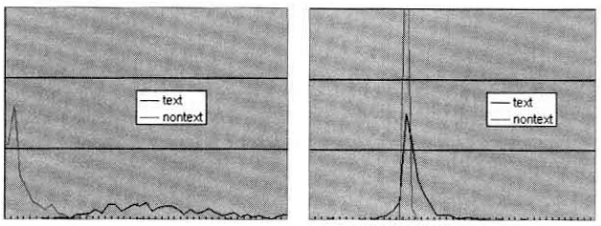
(그림 4) 특징 추출에 사용된 16×16 크기의 텍스트 영상의 예 : (a) 텍스트, (b) 텍스트와 유사한 비텍스트, (c) 비텍스트

16×16 크기의 영상에 반복적으로 웨이블릿 변환을 적용하면 4-레벨 웨이블릿 변환에서는 오직 한 개의 화소만이 남게 되므로 위의 계산은 처음의 세 단계의 부밴드들에서만 계산한다. 따라서 각각의 16×16 크기의 영상에서는 3-레벨 웨이블릿 변환이 수행되어 12개의 밴드들이 생겨나고, 각 밴드에서 3개의 특징이 생겨나므로 전체 36차원의 특징 벡터가 만들어진다. 본 논문에서는 16×16 크기의 비교적 작은 영상을 웨이블릿 변환하므로 기저함수는 필터의 길이가 짧은 하아(Haar) 함수를 사용하였다[7]. (그림 4)(a)는 특징 추출에 사용된 16×16 크기의 텍스트 영상의 예이고, (그림 4)(b)는 텍스트 영상과 비슷한 특징을 갖는 비텍스트 영상의 예이며, (그림 4)(c)는 고주파 성분이 거의 없는 비텍스트 이미지의 예이다.

3.3 특징 선택

추출된 특징 벡터들은 SVM에 입력되어 학습된다. 이 과정에서 특징의 차원이 너무 크면 많은 수의 학습 데이터와 학습 시간이 필요하다. 또한, 두 집단을 제대로 식별하지 못하는 차수의 특징은 오히려 SVM이 최적 분리면을 찾는데 방해요소가 된다. 그러므로 변별력이 낮은 차수의 특징은 제거하여 특징 벡터의 차원을 줄이는 것이 필요하다. 본 연구에서는 특징 벡터의 차원을 줄이기 위해서 베이지안 에러율을 사용하였다.

베이지안 분류 식은 $P(C|x)=\frac{P(C)P(x|C)}{P(x)}$로서 x는 특징 값으로서 벡터일 수도 있으며, C는 분류 클래스이다. 이 식은 특징 x가 나타났을 때 클래스 C에 속할 사후 확률을 구하는 것으로서 클래스 C에 대한 사전확률 $P(C)$, 클래스 C에서 특징 x가 관찰된 조건부 확률 $P(x|C)$와 특징 x가 관찰될 사전 확률 $P(C)$에 의해서 구해진다[10].



(그림 5) 변별력이 높은 특징 차수(좌) 및 변별력이 낮은 특징 차수(우) 예

본 논문에서는 16×16 크기의 텍스트 데이터 1,000개와 비텍스트 데이터 1,000개를 웨이블릿 변환하고 36차원의 특징을 추출한다. 각각의 데이터에 대한 분류 클래스-텍스트 또는 비텍스트-를 알고 있기 때문에 베이지안 분류 방법으로 각 특징에 대한 에러율을 측정할 수 있다. 측정 후에 에러율이 작은 순서대로 12개를 선택한다. (그림 5)는 두 집단이 잘 분리되어 변별력이 높은 특징과 두 집단이 잘 분리되지 않는 변별력이 낮은 특징 분포를 보여준다.

4. SVM 학습 및 검증

SVM에 대한 자세한 내용은 [7,8]에서 참고할 수 있으며, 간단히 설명하면 다음과 같다. SVM은 Vapnik에 의해 제안된 통계적 학습 이론에 기반한 보편적 접근 방법으로서, 경험적 성능뿐만 아니라 고차원 공간에서의 뛰어난 일반화 성능으로 분류 문제의 해결 방법으로 선호되고 있다. SVM은 기존의 통계적 학습 방법들에서 이용되는 경험적 위험 최소화와 다른 구조적 위험 최소화를 이용하여 일반화 오류를 감소시키는 방법을 취하고 있다. 또한, 다계층 퍼셉트론이나 Radial Basis Function 네트워크와 같은 기존의 다른 보편적인 접근 방법들처럼 SVM역시 패턴인식/분류나 비선형 곡선 함수 추정 등의 다양한 일을 효과적으로 수행할 수 있다. 이진 패턴 분류 문제에서 SVM은 학습 데이터의 성격에 따라 선형 분리 가능한 경우, 선형 분리 가능하지 않은 경우, 비선형 특징 공간의 세 가지로 나눈다. SVM의 기본 원리는 선형 분리 가능한 문제에서 출발하며, 선형 분리 가능하다는 것은 두 집단으로 구분되는 초평면이 존재한다는 의미이다. 선형분리 가능하지 않은 경우는 데이터가 분리면의 반대편에 존재하는 경우가 발생하여, 학습 데이터가 선형적으로 분리되지 못하는 경우이며 SVM은 데이터가 원래 제약조건에서 어느 정도 위배되는지의 수준을 측정하여 선형적인 결정평면을 찾아준다. 마지막 비선형 특징 공간의 경우, 앞에서 소개되었던 경우들과는 달리 학습 데이터가 선형 분리면으로 나뉘지 않는 경우가 대부분이다. 이러한 경우 SVM은 입력 벡터를 고차원 특징공간으로 매핑한 뒤, 고차원 공간에서 선형 분리면을 찾아낸다. 이때 커널 함수의 도움을 받아 특징공간에서 입력 벡터의 내적을 쉽게 구하는 방법을 이용한다. 본 절에서는 텍스트 집단과 비텍스트 집단 데이터의 특징을 학습시키면서 파라미터들을 조절하고 최적의 분류기를 구현하는 과정과 구현된 분류기로 텍스트 후보 영역을 검증하는 과정을 설명한다.

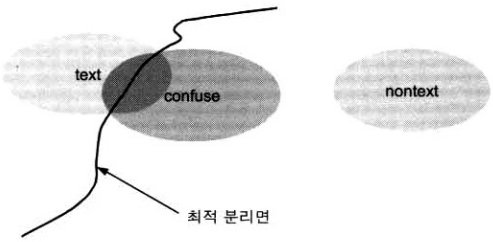
4.1 SVM 학습

SVM 학습에는 텍스트 데이터를 Positive 집단으로, 비텍스트 데이터를 Negative 집단으로 구성하였다. (그림 3)과 같이 비텍스트 집단에는 획 성분이 거의 없는 데이터와 나뭇가지나 나뭇잎 등과 같이 텍스트가 아님에도 불구하고 텍스트와 비슷한 획 성분을 가진 데이터를 포함하고 있다. 후자를 "혼돈(Confuse)" 데이터라고 별도로 정의한다.

텍스트 집단의 데이터 500개를 Positive 집단으로, 획 성분이 거의 없는 비텍스트 집단의 데이터 500개를 Negative 집단으로 학습 집단을 구성하고 각각을 웨이블릿 변환하여

얻은 12차원의 특징 벡터를 입력으로 SVM을 학습하였다. 특징들은 -1에서 1사이의 값을 갖도록 정규화하였다. 학습 결과로 얻어지는 SV들은 텍스트 데이터와 비텍스트 데이터는 잘 분류하는 반면, 대부분의 혼돈 데이터는 텍스트로 분류하는 오류를 발생시킨다. 다음으로 Negative 집단을 비텍스트 데이터 250개와 혼돈 데이터 250개로 구성하고 특징을 추출한 뒤 정규화하여 학습하였다. 그 결과, Negative 집단의 SV는 대부분 혼돈 데이터로 나타났고, 이들은 비텍스트 데이터를 오류없이 잘 분류하였다.

두 번의 학습으로 텍스트 집단, 비텍스트 집단, 혼돈 집단의 구성이 (그림 6)과 같음을 추측할 수 있다. 따라서 텍스트 집단과 혼돈 집단간의 학습만으로도 비텍스트 데이터를 분류할 수 있는 최적 분리면을 얻을 수 있음을 알 수 있다.



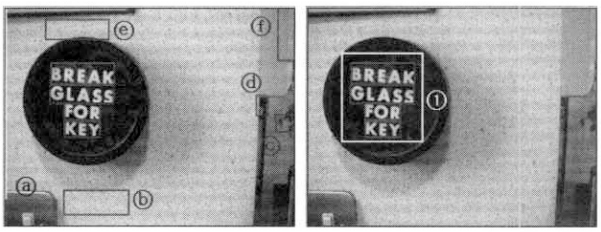
(그림 6) 학습 데이터 집단의 구성

학습 데이터의 수와 Positive, Negative 데이터의 비율, 그 밖의 커널 함수와 그에 따른 파라미터를 결정하기 위해 각각의 수치를 변화시키면서 에러율을 측정한다. 측정결과 학습 집단을 텍스트 데이터 500개와 비텍스트 데이터 500개를 1 : 1의 비율로 하는 것이 가장 높은 정확도를 보였고, σ가 0.2인 가우시안 커널을 사용한 경우에 가장 낮은 에러율을 보였다.

4.2 검 증

텍스트로 추정되는 영역을 검증하기 위해 16×16 크기의 윈도우를 사용하여 검증 영역을 스캔하면서, SVM 분류기를 이용하여 윈도우가 지나가는 위치에 있는 블록의 텍스트 여부를 판별한다. 검증 영역에서 텍스트 블록이 차지하는 비율로써 최종적으로 검증 영역의 텍스트 영역 여부를 결정한다.

우선 입력 이미지가 주어지면 에지 분포를 이용하여 텍스트 후보 영역을 추출하고 후보 영역 중에서 검증해야 할 텍스트 후보 영역의 개수와 좌표를 입력받는다. 입력받은 영역에 16×16 크기의 윈도우를 씌우고 웨이블릿 변환하여 12차원의 특징 벡터를 추출한다. 추출한 특징을 SVM 학습에서 구한 SV들과의 거리를 계산하여 윈도우 안의 블록이 텍스트 블록인가 아닌가를 판별한다. 블록의 텍스트 여부를 판단한 뒤, 4화소를 이동한 위치에서 이와 같은 블록의 판별을 반복한다. 이 때, 이동하는 화소의 크기를 작게 할수록 블록 판별의 정확도를 높일 수 있으나 판별해야하는 블록의 개수가 증가하므로 검증속도는 저하된다. 영역에 대한 수행이 끝나면, 한 영역에서 생겨난 모든 블록 중 텍스트로 판별된 블록의 수를 세고, 텍스트로 판별된 블록의 수가 전체 블록의 25% 이상이면 영역을 텍스트 영역으로 판단하고 그렇지 않은 경우 영역을 비텍스트 영역으로 결정한다. 이와 같은 처리과정을 모든 텍스트 후보영역에 대하여 수행한다.



(그림 7) 텍스트 후보 영역(좌) 및 SVM 검증 결과(우)

(그림 7)은 텍스트 후보영역과 이를 검증한 결과이다. (그림 7)(좌)에서 사각형 박스로 표시된 영역은 이미지의 에지 분포를 이용하여 추출한 텍스트 후보 영역이다[11]. 그 중 ⓐ~ⓕ 영역은 획 성분이 거의 없거나 또는 획 성분의 조합이 텍스트와 달라 다해상도 웨이블릿을 이용한 특징 추출 단계에서 SV와 거리가 먼 특징 값이 추출되었으므로 검증 과정에서 제거되었다. 영역 ①은 추출된 특징 값이 SV와 유사하여 검증 과정에서 텍스트 영역으로 판단되었다.

이미지에 포함된 텍스트의 크기가 (그림 8)과 같이 16×16보다 훨씬 큰 경우 오류가 발생한다. 본 연구에서는 이와 같이 글자 크기로 인해 발생되는 오류를 줄이기 위해서 이미지를 가로와 세로를 각각 1/2로 축소하고 높이와 너비가 16화소 이상인 박스 영역에 대해서 같은 방법으로 검증을 수행한 뒤, 320×240 크기의 이미지에서의 검증 결과와 OR 결합을 수행한다. 이미지를 축소함으로써 이미지에 포함된 크기가 큰 문자는 16×16 윈도우에 적합한 글자 크기를 바꾸게 되어 검증 단계에서 텍스트로 분류될 수 있었다. (그림 9)는 글자의 크기가 커서 오류가 발생한 320×240 크기의 이미지를 160×120 이미지로 축소한 뒤 검증을 수행하여 텍스트로 분류한 결과이다.

(그림 8) 이미지에서의 글자 크기로 인한 오류 발생 예

(a) (b) (c) (d)

(그림 9) (a) 에지 분포를 이용한 후보영역 추출, (b) 320×240 크기의 이미지에서 검증 수행, (c) 160×120 크기의 이미지에서 검증 수행, (d), (b)와 (c)의 OR 결합 결과

5. 실험 및 결과

본 연구에서 제안한 방법은 Windows XP에서 Visual C++ 6.0을 이용하여 Pentium IV 1.8GHz 하드웨어 상에서 구현하였다. 실험은 이미지의 에지 분포를 이용하여 추출한 후보 영역을 대상으로 하여 SVM 검증 과정을 추가하였을 때와 추가하지 않았을 때의 결과를 비교하였다. 이미지의 에지 분포를 이용한 문자 추출 방법은 입력된 이미지의 에지 이미지를 구하고, 에지 분포를 분석하여 이미지에 포함되어 있는 다양한 모양의 긴 선들을 제거하여 처리할 이미지를 단순하게 만든 뒤, 모폴로지를 적용하여 각 연결요소를 강조한 후 연결요소를 분석, 검증하여 최종적으로 후보 영역을 추출하는 방법이다[11].

실험에 사용된 이미지는 학교, 병원, 지하철 역, 도로 등의 실내외에서 디지털 카메라로 취득한 120개의 자연영상과 실외 간판영상 100개, ICDAR(International Conference on Document Analysis and Recognition)에서 컨테스트용으로 제공한 100개의 영상을 사용하였다. 자연영상은 배경의 복잡도에 따라 단순한 이미지와 복잡한 이미지 그룹으로 분류하여 실험하였다.

실험 결과는 일반적으로 사용되는 정확률(precision), 재현율(recall)과 이를 조합한 채산점(precision/recall break-even point)으로 평가하였다. 각각을 <표 1>로부터 이끌어 낸다. 텍스트를 비텍스트로 분류하면 양의 오류(false positive)로서 a로 표현하며, 비텍스트를 텍스트로 분류하면 음의 오류(false negative)로서 b로 표현한다. 따라서 정확률은 텍스트로 분류된 것 중에서 정확히 텍스트의 비율을 표현하는 것으로서 비텍스트가 텍스트로 분류되는 오류가 많아져도 정확률은 낮아진다. 재현율은 비텍스트가 텍스트로 분류된 것은 상관하지 않으며 단지 텍스트가 몇 개 찾아졌는지에 대한 평가를 내리는 것이다. 따라서 응용에 따라 정확률 또는 재현율이 높은 것을 선택하여 사용하는 것이 바람직하다.

〈표 1〉 오류의 유형

	텍스트	비텍스트
텍스트로 분류	a	b
비텍스트로 분류	c	d

$$Precision = \frac{a}{a+b}$$

$$Recall = \frac{a}{a+c}$$

$$(precision/recall\ break-even\ point)$$

$$= \frac{2 \times precision \times recall}{precision + recall}$$

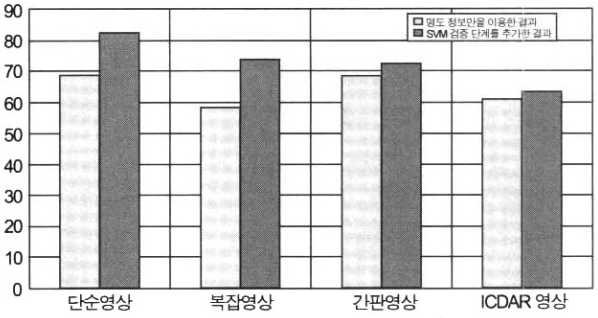
이미지의 에지 분포를 이용하여 추출된 텍스트 영역의 결과는 <표 2>와 같으며, <표 3>은 에지 정보로 추출된 텍스트 후보영역을 SVM 분류기로 검증한 결과이다. *Total*은 실험 대상 이미지에서의 전체 텍스트 수이며, *Correct*는 정확히 찾은 텍스트 개수, *Missing*은 텍스트를 찾지 못한 개수이며, *False*는 텍스트가 아닌 영역을 텍스트로 찾은 개수이다.

〈표 2〉 이미지의 에지 분포를 이용한 추출 결과

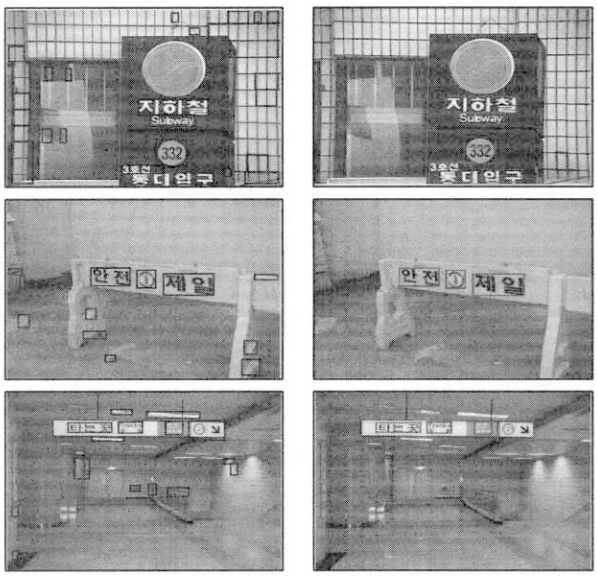
			Total	Correct	Partial	Missing	False
자연영상	단순 이미지	영역 개수	256	228	3	28	177
		Precision = **56.2%**		Recall = **89.0%**		채산점 = **68.9%**	
	복잡 이미지	영역 개수	328	233	17	95	234
		Precision = **49.8%**		Recall = **71.0%**		채산점 = **58.5%**	
간판영상		영역 개수	839	632	40	167	387
		Precision = **63%**		Recall = **75%**		채산점 = **68.5%**	
ICDAR 컨테스트 영상		영역 개수	687	455	19	213	354
		Precision = **57%**		Recall = **66%**		채산점 = **61.2%**	

〈표 3〉 SVM 검증 후

			Total	Correct	Partial	Missing	False
자연영상	단순 이미지	영역 개수	256	222	3	34	54
		Precision = **78.7%**		Recall = **86.7%**		채산점 = **82.5%**	
	복잡 이미지	영역 개수	328	224	17	104	87
		Precision = **79.1%**		Recall = **69.1%**		채산점 = **73.8%**	
간판영상		영역 개수	839	602	40	197	180
		Precision = **73.2%**		Recall = **71.8%**		채산점 = **72.5%**	
ICDAR 컨테스트 영상		영역 개수	687	409	19	259	173
		Precision = **67.9%**		Recall = **59.5%**		채산점 = **63.4%**	



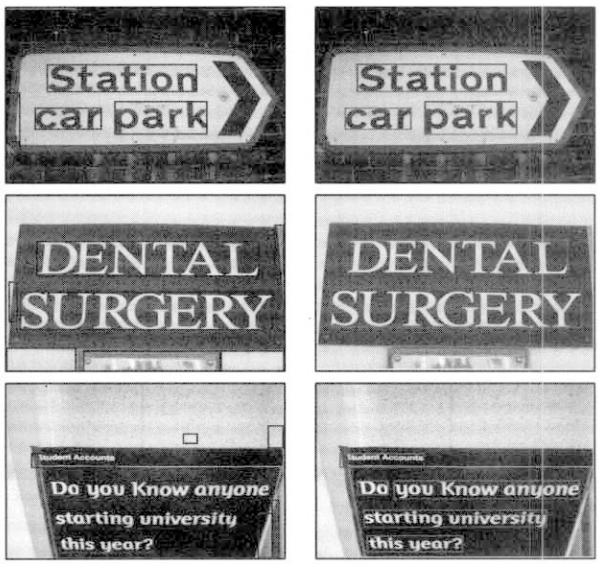
(그림 10) 검증 전과 SVM 검증 후의 채산점 비교



(그림 11) 단순한 이미지에 대한 후보영역 추출(좌) 및 SVM 검증(우)

(그림 12) 복잡한 이미지에 대한 후보영역 추출(좌) 및 SVM 검증(우)

(그림 13) 간판영상에 대한 후보영역 추출(좌) 및 SVM 검증(우)



(그림 14) ICDAR 컨테스트 영상에 대한 후보영역 추출(좌) 및 SVM 검증(우)

<표 2>에 비해 <표 3>의 재현율은 다소 감소하였고 정확률은 대폭 증가하였다. SVM 검증이 이미지의 에지 분포를 이용하여 추출된 후보영역만을 대상으로 하였기 때문에 *Correct*는 더 이상 증가시킬 수 없고, *Missing*만이 더 생겨날 수 있으므로 재현율의 수치는 동일하거나 감소할 수 밖에 없다. 그러나 이미지의 에지 분포만을 이용하였을 때는 잘 분류할 수 없었던 건물벽면의 타일, 나뭇가지, 나뭇잎과 같이 텍스트와 유사한 특징을 갖는 영역을 SVM 검증을 통해 제거해서 *False*의 수치를 줄일 수 있었기 때문에 정확률은 월등히 향상되었다. 특히, 복잡한 이미지는 텍스트와 유사한 특징을 갖는 영역을 많이 포함하므로 복잡한 이미지의 경우 정확률의 상승효과가 컸다. (그림 10)은 재현율의 감소보다 정확률의 증가가 더 컸기 때문에 채산도가 전반적으로 증가함을 보여준다. (그림 11), (그림 12), (그림 13), (그림 14)는 각각 자연영상, 간판영상, ICDAR 컨테스트 영상에 대한 SVM 검증 결과를 보여준다. 각 그림에서 왼쪽 열의 그림은 에지를 이용하여 후보영역을 추출한 결과이며, 오른쪽 열은 그 영역들을 SVM으로 검증한 결과이다.

6. 결 론

최근 패턴인식 분야에서 주목받고 있는 SVM과 웨이블릿 변환을 이용하여 문자 영역을 검증하는 방법을 제안하였다. 제안한 방법을 적용함으로써 벽면의 타일이나 나뭇가지, 나뭇잎 등과 같이 텍스트와 유사한 특성을 갖는 영역을 비교적 잘 분류해 낼 수 있었다. 또한, 텍스트의 크기가 큰 경우 영상의 크기를 축소하여 검증함으로써 문자 크기로 인해서 발생하는 오류를 줄일 수 있었다. 향후 연구로는 본 연구 결과를 일반적인 텍스트 영역의 추출 방법과 결합하여 응용시스템을 구축하는 것이며, 자연이미지에서의 일반적인 텍스트 영역의 추출 방법에 대한 최근 결과 및 방향은 [13, 14]를 참고할 수 있다.

참 고 문 헌

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

[2] Yu Zhong, Kalle Karu, Anil K. Jain, "Locating Text in Complex Images," Pattern Recognition, Vol.28, No.10, pp. 1523-1535, 1995.

[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] Huiping Li, David Doermann and Omid Kia, "Automatic Text Detection and Tracking in Digital Video," IEEE Transactions on Image Processing, Vol.9, No.1, pp.147-156, 2000.

[5] K. I. Kim, "Texture-Based Approach For Text Detection In Images Using Support Vector Machines and Continuously Adaptive Mean Shift Algorithm," IEEE Trans. Pattern Analysis and Machine Intelligence(TPAMI), to be published.

[6] Ki-Young Jeong, Keechul Jung, Hang Joon Kim, "Neural Network-Based Text Location for News Video Indexing," International Conference on Image Processing(ICIP), 3, pp.319-323, 1999.

[7] Scott E Umbaugh, Computer Vision and Image Processing, Prentice Hall PTR, New Jersey, 1999.

[8] V. N. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.

[9] C. Cortes, V. Vapnik, "Support Vector Networks," In Proceedings of Machine Learning, Vol.20, pp. 273-297, 1995.

[10] E. Gose, R. Johnsonbaugh, S. Jost, Pattern Recognition and Image Analysis, Prentice Hall PTR, 1996.

[11] 김길천, 최영우, 변혜란, "명도 정보를 이용한 장면 텍스트 추출", 한국정보과학회 컴퓨터비전 및 패턴인식연구회 추계워크샵 발표논문집, 서울, pp.159-160, 2001.

[12] Yu Zhong, Kalle Karu, Anil K. Jain, "Locating Text in Complex Images," Pattern Recognition, Vol.28, No.10, pp. 1523-1535, 1995.

[13] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong and R. Young, "ICDAR 2003 Robust Reading Competition," ICDAR 2003, Vol.2, pp.682-687, 2003.

[14] David Doermann, Jian Liang and Huiping Li, "Progress in Camera-Based Document Image Analysis," ICDAR 2003, Vol.1, pp.606-616, 2003.

배 경 숙

e-mail : pavin@etri.re.kr

1998년~2002년 숙명여자대학교(이학사)

2002년~2004년 숙명여자대학교(이학석사)

2004년~현재 한국전자통신연구원 지능형 로봇연구단 연구원

관심분야 : 영상처리, 문자추출, 패턴인식 등

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 2월 LG전자기술원 선임연구원

1997년~현재 숙명여자대학교 정보과학부 조교수, 부교수

관심분야 : 영상처리, 패턴인식, 문자인식 등