

논문 2017-54-2-15

시각장애인을 위한 딥러닝 기반 표지판 검출 및 인식

(Deep Learning Based Sign Detection and Recognition for the Blind)

전 태 재*, 이 상 윤**

(Taejae Jeon and Sangyoun Lee[Ⓢ])

요 약

본 논문은 딥러닝 알고리즘을 기반으로 하여 시각장애인을 위한 표지판을 검출하고 인식하는 시스템을 제안한다. 제안된 시스템은 크게 표지판 검출 단계와 표지판 인식 단계로 나뉘는데 표지판 검출 단계에서는 영상에서 응집 채널 특징을 추출한 뒤 아다부스트 분류기를 적용하여 표지판 관심영역을 검출하였고, 표지판 인식 단계에서는 검출한 표지판 관심영역들에 합성곱 신경망을 적용하여 어떤 표지판인지 인식하였다. 본 논문에서는 미검출된 표지판의 개수가 최대한 감소하도록 아다부스트 분류기를 설계하였고, 딥러닝 알고리즘을 사용하여 인식 정확도를 높임으로써 검출 단계에서 발생한 양성 오류들을 제거시켰다. 실험 결과, 제안된 방법의 양성 오류 개수가 다른 방법들의 양성 오류 개수보다 효과적으로 감소했음을 확인하였다.

Abstract

This paper proposes a deep learning algorithm based sign detection and recognition system for the blind. The proposed system is composed of sign detection stage and sign recognition stage. In the sign detection stage, aggregated channel features are extracted and AdaBoost classifier is applied to detect regions of interest of the sign. In the sign recognition stage, convolutional neural network is applied to recognize the regions of interest of the sign. In this paper, the AdaBoost classifier is designed to decrease the number of undetected signs, and deep learning algorithm is used to increase recognition accuracy and which leads to removing false positives which occur in the sign detection stage. Based on our experiments, proposed method efficiently decreases the number of false positives compared with other methods.

Keywords : deep learning, sign detection and recognition, blind assistance system, AdaBoost classifier, convolutional neural network

I. 서 론

오늘날 영상인식 기술이 지속적으로 발전하면서 시각장애인들의 보행을 보조하는 기술이 연구되고 있다. 시각장애인들의 불편한 눈을 대신하기 위해 카메라로 받아들인 영상을 영상인식 기술로 처리하여 시각장애인들에게 도움이 될 수 있는 정보를 제공하는 것이다. 이러한 정보 제공의 대표적인 예로 장애물 검출을 통한

장애물 회피, 신호등 검출 및 인식을 통한 횡단보도 횡단, 표지판 검출 및 인식을 통한 주요 시설 위치 제공 등이 있다. 그 중에서 시각장애인을 위한 표지판 검출 및 인식 분야는 아직 개발 수준이 높지 않기 때문에 꾸준한 연구를 필요로 하는 분야이다. 하지만 공개된 시각장애인을 위한 표지판 데이터베이스가 없었기 때문에 자체제작한 표지판 데이터베이스를 사용하였다. 남자 화장실, 비상구, 지하철, 장애인, 계단 표지판 영상에 표지판이 없는 배경 영상을 더하여 총 6개 클래스의 영상들로 구성되어 있다. 또한 앞서 언급했듯이 관련 연구가 많이 없었기 때문에 표지판의 종류와 활용도만 다를 뿐 유사한 연구 분야인 교통표지판 검출 및 인식 분야를 바탕으로 연구를 수행하였다.

교통표지판 검출 및 인식에 관한 기존의 연구 방법들 중에 많은 비중을 차지하는 방법 중 하나가 특정 색상

* 학생회원, ** 평생회원, 연세대학교 전기전자공학과 (School of Electrical and Electronic Engineering, Yonsei University)

Ⓢ Corresponding Author (E-mail : syleee@yonsei.ac.kr)

※ 본 연구는 국토해양부 및 한국건설교통기술평가원 교통물류연구사업(16TLRP-B079209-03, 교통약자 보행지원 시스템 개발)의 연구비지원에 의해 수행되었음.

Received ; November 21, 2016 Revised ; December 18, 2016

Accepted ; January 24, 2017

을 추출한 뒤 활용하는 방법이다. 이 방법은 교통표지판의 테두리나 전체 색상이 빨간색, 파란색이라는 정보를 이용하여 빨간색, 파란색의 색상 영역을 추출한 뒤 해당 영역에서 검출 및 인식을 수행하는 방법이다. 관련 연구로는 교통표지판의 색 정보와 수직면 특징을 이용하여 표지판 후보군을 검출한 뒤 교통표지판 패턴에 부합하는 다양한 특징 원형들과 아다부스트 분류기(AdaBoost classifier)를 통해 검출 및 인식하는 방법^[1], 교통표지판 영상들로부터 생성한 HSI 고유칼라 모델을 이용하여 표지판 후보 영역을 검출한 뒤 불변 모멘트 특징 정보와 서포트 벡터 머신(Support Vector Machine)을 통해 검출 및 인식하는 방법^[2], 교통표지판의 색상과 색상 윤곽선을 추출한 뒤 투표 기법을 이용하여 교통표지판의 형태를 계산하여 스코어 맵들(score maps)을 생성한 후 이 스코어 맵들을 이용하여 서포트 벡터 머신을 학습하는 방법^[3] 등이 있다. 하지만 시각장애인을 위한 표지판의 경우 이 방법을 적용하기가 어려운데 그 이유는 교통표지판의 경우 국가에서 지정한 색상으로 통일되어 사용되지만 시각장애인을 위한 표지판의 경우 주로 사용되는 색상이 있지만 다른 색상을 사용하는 경우도 있기 때문이다.

한편, 최근 교통표지판 검출 및 인식 연구에 많이 이용되는 방법으로 딥러닝(deep learning)을 이용하는 방법이 있다. 딥러닝은 여러 비선형 변환기법의 조합을 통해 높은 수준의 추상화(abstractions)를 시도하는 기계학습(machine learning) 알고리즘의 집합^[4]으로 정의되는데, 다양한 영상인식 알고리즘에서 높은 성능을 보여 최근에 각광받고 있는 기술이다. 특히, 딥러닝 알고리즘 중에 영상인식에 많이 사용되는 알고리즘으로 합성곱 신경망(Convolutional Neural Network) 알고리즘이 있는데 학습 영상의 특징을 잘 잡아내는 장점이 있다. 관련 연구로는 합성곱 신경망 구조 하나에서 검출과 인식을 동시에 수행하도록 설계한 방법^[5], 입력 영상으로 이미지 피라미드(image pyramid)를 생성한 뒤 슬라이딩 윈도우(sliding window) 방식으로 합성곱 신경망을 적용하여 검출 및 인식을 수행하는 방법^[6] 등이 있다. 합성곱 신경망 구조 하나에서 검출, 인식을 동시에 수행할 경우 검출의 정확도가 떨어지는 단점이 있으며 슬라이딩 윈도우 방식으로 합성곱 신경망을 적용할 경우 알고리즘 속도 때문에 간단한 신경망 구조를 사용하게 되는데 이에 따른 인식 성능 저하의 문제가 있다. 본 논문에서는 딥러닝 알고리즘을 이용하여 시각장애인을 위한 표지판을 검출, 인식하는 시스템을 제안한다.

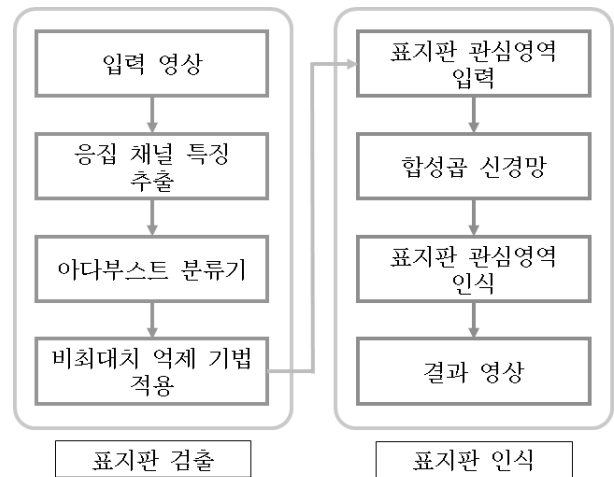


그림 1. 제안된 시스템의 흐름도

Fig. 1. The flowchart of the proposed system.

제안된 시스템은 딥러닝 알고리즘을 표지판 검출 부분에서는 사용하지 않고 인식 부분에서만 사용하였는데, 그 이유는 딥러닝 검출 알고리즘의 성능이 다른 알고리즘보다 떨어지기 때문이다. 자체제작한 표지판 데이터베이스의 경우 표지판의 크기가 매우 큰 것부터 매우 작은 것까지 다양한데 딥러닝 검출 알고리즘의 경우 검출 가능한 물체의 크기에 제약이 있으며, 물체의 위치를 정확히 찾기 위해서는 복잡한 구조의 신경망을 사용해야 하는데 이 경우 알고리즘 동작 속도가 느려지게 된다. 따라서 표지판 검출 부분에는 비교적 가볍고 성능이 좋은 아다부스트 분류기를 사용하였고 높은 성능을 보이는 인식 부분에 딥러닝 알고리즘을 적용하였다.

그림 1은 본 논문에서 제안된 시스템의 흐름도이다. 먼저 표지판 검출 부분에서는 입력 영상에서 응집 채널 특징을 추출하고 트리를 이용한 아다부스트 분류기를 적용하여 표지판 관심영역(Region Of Interest)을 찾은 뒤 비최대치 억제 기법(non-maximum suppression)을 적용하여 중복되는 영역들을 제거한다. 그 후, 표지판 인식 부분에서는 검출 부분에서 찾은 관심영역들에 딥러닝 알고리즘 중에 하나인 합성곱 신경망을 적용하여 각각의 관심영역이 어떤 표지판인지를 인식한 결과를 얻게 되고 이 결과들을 종합하여 최종 결과 영상을 출력하게 된다.

본 논문의 구성은 다음과 같다. II장에서 아다부스트 분류기를 이용한 표지판 검출 알고리즘을 기술하고, III장에서 합성곱 신경망을 이용한 표지판 인식 알고리즘을 설명한다. 그리고 IV장에서 실험 결과와 이에 대한 분석을 서술하며, 마지막으로 V장에서 본 논문의 결론을 맺는다.

II. 아다부스트 분류기를 이용한 표지판 검출

1. 응집 채널 특징

먼저 그림 2의 예시와 같이 입력 영상에서 채널 영상들을 계산한다. 이 때 사용되는 채널은 방향별 윤곽선(edge) 채널 6개, 윤곽선 채널 1개, LUV 색상채널 3개이다. 각 채널 영상을 구한 뒤 영상을 4×4 크기의 사각형 블록들로 나눈 후 각 블록 내의 픽셀 값들을 모두 더해 특징 맵(feature map)을 생성한다. 이렇게 생성된 특징 맵은 원래 채널영상보다 1/16배만큼 크기가 작아진다. 그리고 이 특징 맵의 각 픽셀 값들을 일렬로 이어서 특징벡터를 생성하는데 이 특징 벡터를 응집 채널 특징이라 한다^[7]. 이 특징을 사용하여 뒷단에서 깊이가 2인 포화 이진 트리를 학습하게 된다. 루트 노드(root node)와 그 하위 노드에서 사용하는 분류 기준은 앞서 구한 특징벡터에서 무작위 위치의 값을 선택하여 긍정 표본(positive sample)과 부정 표본(negative sample)을 가장 잘 나누는 임계값을 기준으로 하였다. 모델의 크기는 64×64 , 128×128 의 두 개의 크기를 사용하였다. 두 개의 모델을 학습한 이유는 모델의 크기가 한 개일 경우 다양한 크기의 물체에 제대로 매칭이 되기 어렵고 영상의 크기를 계속 변경해가며 슬라이딩 윈도우 기법을 적용해야하기 때문에 실험 시간이 오래 걸린다는 단점이 있다. 하지만 모델의 개수가 너무 많을 경우에는 학습하는데 시간이 오래 걸린다는 단점이 있다. 따라서 절충점을 찾아 두 개의 모델 크기를 사용하였다.

2. 아다부스트 분류기

아다부스트의 약한 분류기로 깊이가 2인 포화 이진 트리를 사용한다. 포화 이진 트리는 모든 단말 노드의 깊이가 같은 정 이진 트리로서 깊이가 2라는 것은 루트 노드에서 단말 노드까지 가는 경로의 길이가 2라는 것을 의미한다. 깊이가 2인 트리를 사용한 이유는 깊이가 1인 트리를 약한 분류기로 사용했을 경우 분류기가 너무 단순한 구조이기 때문에 분류력이 떨어진다고 판단했기 때문이다. 아다부스트 알고리즘은 약한 분류기를 여러 개 합쳐서 강한 분류기를 만드는데 깊이가 2인 포화 이진 트리는 이에 적합한 형태의 약한 분류기이다. 아다부스트 분류기를 학습하는 과정에서 가중치가 2개 사용된다. 하나는 각각의 학습 데이터들에 적용되는 가중치이고 나머지 하나는 각각의 약한 분류기들에 적용되는 가중치이다. 하나의 약한 분류기를 학습하고 나면 그 약한 분류기의 성능을 테스트해서 성능이 좋으면 가

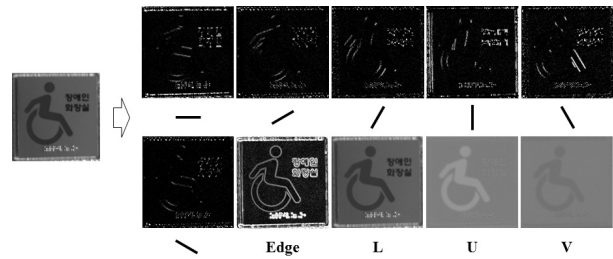


그림 2. 채널 영상의 예시
Fig. 2. The example of computed channel images.

중치를 높게 설정하고 성능이 나쁘면 가중치를 낮게 설정한다. 이로 인해 나중에 강한 분류기를 만들 때 분류 성능이 높은 분류기의 가중치가 커서 분류를 하는데 있어서 큰 영향력을 보일 수 있게 된다. 그리고 해당 약한 분류기까지 학습된 모든 약한 분류기들로 학습 데이터를 다시 테스트해서 분류가 잘 되는 데이터는 가중치를 낮게 설정하고 분류가 잘 안 되는 데이터는 가중치를 높게 설정해서 다음 분류기를 학습할 때 잘못 분류된 데이터가 더 잘 학습되도록 유도한다. 이런 식으로 약한 분류기를 학습하는 것을 반복하고 원하는 개수의 약한 분류기가 학습되면 이 약한 분류기들에 학습하면서 조정된 가중치를 적용해 합쳐서 강한 분류기를 만들게 된다^[8]. 정확도 향상을 위해 이러한 강한 분류기를 연달아 4개를 학습하고 실험할 때에는 4개의 강한 분류기에서 나온 결과의 평균을 사용하는 방식을 썼다. 그리고 연산 속도 향상을 위해 4개의 강한 분류기 중에 하나라도 실험 영상을 부정 표본으로 판단할 경우 최종적으로 부정 표본이라고 확정하고 그 이후의 연산은 하지 않는 방식을 적용하였다. 강한 분류기를 학습할 때는 같은 조건에서 학습하지 않고 강한 분류기를 하나 학습할 때마다 약한 분류기의 개수를 늘리고 부정 표본, 강한 부정 표본(hard negative sample)을 추가하는 등의 변화를 주었다. 여기에서 강한 부정 표본이란 이전까지 학습된 모든 강한 분류기를 통해 실험했을 때 긍정 표본으로 잘못 분류된 부정 표본을 말하는데, 부정 표본이지만 긍정 표본의 특징을 갖고 있어서 분류하기 어려운 부정 표본이다. 이러한 강한 부정 표본을 일정 비율만큼 추가하여 학습할 경우 분류하기 어려운 부정 표본을 더 잘 분류할 수 있게 되기 때문에 양성 오류(false positive) 개수가 감소하게 된다. 표 1은 실험할 때 설정했던 약한 분류기와 부정 표본, 강한 부정 표본의 개수이다. 세 가지 항목의 개수를 증가시키면서 학습한 이유는 다음 강한 분류기를 학습할 때 이전에 학습했던 강한 분류기보다 분류 성능을 더 높이고 양성

표 1. 학습에 사용했던 약한 분류기, 부정 표본, 강한 부정 표본의 개수

Table1. The number of weak classifiers, negative samples and hard negative samples used in training.

| | 1번째 강한 분류기 | 2번째 강한 분류기 | 3번째 강한 분류기 | 4번째 강한 분류기 |
|----------|------------|------------|------------|------------|
| 약한 분류기 | 50 | 50 | 100 | 200 |
| 부정 표본 | 2000 | 3900 | 5820 | 7760 |
| 강한 부정 표본 | 0 | 100 | 180 | 240 |

오류에 더 강한 모델을 학습하기 위함이다. 긍정 표본의 경우 각 클래스의 학습 영상들을 변형이나 추가 없이 그대로 사용했다. 위의 방법으로 5개의 표지판에 대한 분류기를 각각 생성하였다.

3. 비최대치 억제 기법

비최대치 억제 기법은 하나의 물체 주위에 여러 개의 윈도우가 검출됐을 경우 가장 그 물체라고 판단되는 윈도우 하나를 제외한 나머지 윈도우들을 제거하는 기법이다. 모든 윈도우들을 두 개씩 묶어서 두 윈도우가 서로 겹치는 영역의 넓이를 두 윈도우의 영역 중에 작은 영역의 넓이로 나눈 값이 0.4보다 클 경우 두 윈도우가 같은 물체를 찾은 것으로 간주하여 아다부스트 분류기의 신뢰도 값이 큰 윈도우를 남기고 나머지 윈도우를 제거하였다. 이런 기준을 세운 이유는 표지판의 크기가 크게 촬영된 영상에서 표지판 내부에 발생한 윈도우들을 제거하기 위함이다. 이 기법을 사용함으로써 물체가 하나 있음에도 불구하고 그 주위에 복수 개의 윈도우가

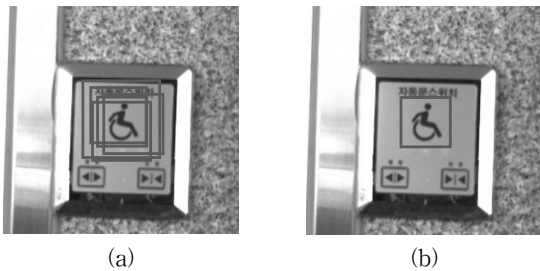


그림 3. (a) 비최대치 억제 기법 적용 전 영상 (b) 비최대치 억제 기법 적용 후 영상

Fig. 3. (a) An image before applying non-maximum suppression (b) An image after applying non-maximum suppression.

검출되어 발생하는 양성 오류의 개수를 감소시킬 수 있다. 비최대치 억제 기법의 예시는 그림 3과 같다.

Ⅲ. 합성곱 신경망을 이용한 표지판 인식

1. 영상 크로핑

표지판 인식기를 학습하기 위해 학습용 데이터베이스에 영상 크로핑을 수행하였다. 영상 크로핑을 하는 목적은 학습 시 정보가 담겨있지 않은 불필요한 영역을 제거하여 인식률을 향상시키기 위함이다. 원본 영상은 표지판 영상이고, 여기에서 인식에 영향을 주는 주요 부분을 잘라내어 크로핑된 영상을 얻는다. 마지막으로 영상의 크기를 48×48로 조절하여 정사각형이 되게 하였다. 이렇게 크기가 조절된 영상을 학습용 및 실험용 영상으로 사용하였다. 그림 4는 영상 크로핑의 예를 보여준다.

2. 합성곱 신경망

표지판 인식에 사용한 합성곱 신경망은 심층 인공 신경망(Deep Neural Network) 중 하나로서 9개의 레이어(layer)로 구성되어 있는데 표 2와 같은 구조를 가지고 있다^[9]. 입력(input) 레이어, 컨볼루션(convolutional) 레이어, 맥스 풀링(max pooling) 레이어, 풀리 커넥티드(fully connected) 레이어로 이루어진 계층 구조를 띄고 있는데 여기에서 컨볼루션 레이어가 사용되기 때문에 합성곱 신경망이라 불린다. 먼저 레이어 0은 영상을 입력 받는 입력 레이어이다. 표지판 데이터베이스가 색상 정보가 있는 영상이기 때문에 입력 레이어에서 빨강, 초록, 파랑의 3개의 맵(map)을 형성하며 각 맵에서 뉴런(neuron)의 개수는 영상의 픽셀 수와 동일하다. 표지판 데이터베이스에서 영상 크로핑을 통해 48×48 크기의 영상이 생성되었기 때문에 48×48개의 뉴런으로 한 개의 맵이 구성된다. 레이어 1, 레이어 3, 레이어 5는 앞

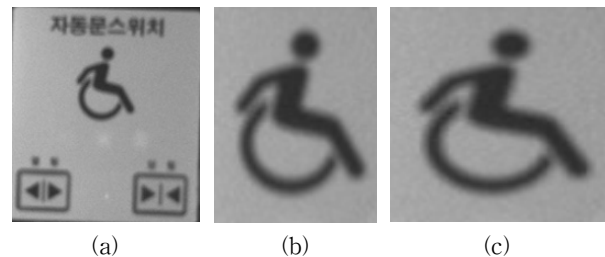


그림 4. (a) 원본 영상, (b) 크로핑 영상, (c) 48×48 영상
Fig. 4. (a) Original image, (b) Cropped image, (c) An image resized to 48×48.

표 2. 합성곱 신경망 알고리즘의 구조
Table2. The architecture of convolutional neural network.

| 레이어 | 종 류 | 맵과 뉴런의 개수 | 커널 크기 | 비율 |
|-----|-------------|---------------------|-------|-----|
| 0 | 입력 레이어 | 48 × 48개 뉴런의 맵 3개 | | |
| 1 | 컨볼루션 레이어 | 42 × 42개 뉴런의 맵 100개 | 7 × 7 | |
| 2 | 맥스 풀링 레이어 | 21 × 21개 뉴런의 맵 150개 | 2 × 2 | |
| 3 | 컨볼루션 레이어 | 18 × 18개 뉴런의 맵 150개 | 4 × 4 | |
| 4 | 맥스 풀링 레이어 | 9 × 9개 뉴런의 맵 150개 | 2 × 2 | |
| 5 | 컨볼루션 레이어 | 6 × 6개 뉴런의 맵 250개 | 4 × 4 | |
| 6 | 맥스 풀링 레이어 | 3 × 3개 뉴런의 맵 250개 | 2 × 2 | |
| 7 | 풀리 커넥티드 레이어 | 300개 뉴런 | 1 × 1 | |
| | 드롭아웃 | | | 0.8 |
| 8 | 풀리 커넥티드 레이어 | 6개 뉴런 | 1 × 1 | |

의 레이어의 맵에서 컨볼루션 연산을 수행하여 새로운 특징 맵을 형성하는 단계이며 컨볼루션 레이어의 맵의 개수는 심층 인공 신경망의 히든(hidden) 레이어에서의 노드(node)의 개수를 뜻한다. 특징 맵 개수는 컨볼루션 레이어를 거칠수록 점점 증가하기 때문에 레이어가 깊어질수록 자세한 특징을 점점 더 많이 나타내게 된다. 또한 컨볼루션 레이어를 거칠 때마다 특징 맵을 구성하는 뉴런의 개수가 줄어들게 되는데 그 이유는 컨볼루션을 수행할 때 특징 맵의 바깥 부분에서는 온전히 값을 구할 수 없기 때문에 값을 버리기 때문이다. 레이어 2, 레이어 4, 레이어 6은 맥스 풀링 레이어로 앞의 컨볼루션 레이어에서 얻은 특징 맵의 크기를 줄임으로써 변형에 강인한 특성을 지니게 되며, 커널(kernel)의 크기는 2 × 2를 사용하였다. 맥스 풀링 레이어의 커널 크기는 일반적으로 영상의 가로, 세로 크기가 100 픽셀 이하일 때 2 × 2 크기를 많이 사용하며 그 이상을 넘을 때 더 큰 커널 크기를 사용하기 때문에 본 신경망 구조에서는 2 × 2 크기를 사용하였다. 레이어 7, 레이어 8은 풀리 커넥티드 레이어로 레이어 6의 특징 맵들의 뉴런들의 값들을 기하학적 요소가 없는 하나의 뉴런 벡터의 각 요소들에 모두 대응시키게 된다. 두 개의 풀리 커넥티드 레이어를 사용함으로써 갑작스러운 뉴런 개수의 감소로 인한 정보 손실을 막게 되며, 마지막 레이어에서는 표지판 클래스의 개수 5개와 배경을 나타내는 클레

스 1개를 합친 총 6개의 클래스에 대응하는 6개의 뉴런들의 값이 출력된다. 각 뉴런들의 값은 0에서 1 사이이며 그 중에 가장 큰 값을 출력하는 클래스를 해당 입력 영상의 클래스로 판단하게 된다. 배경 영상을 하나의 클래스로 설정한 이유는 앞서 표지판 검출 단계에서 표지판과 유사한 배경으로 인해 발생한 양성 오류의 경우 높은 성능을 보여주는 합성곱 신경망의 인식기를 사용하면 배경 클래스로 구분되기 때문에 제거 가능성이 높아지기 때문이다. 그리고 레이어 7의 풀리 커넥티드 레이어를 학습할 때는 드롭아웃(dropout)이라는 정규화 기법을 사용하였다. 학습 시 십만 번 이상의 시행(iteration)을 반복하게 되는데 드롭아웃 정규화를 사용할 경우 각 시행마다 레이어 내의 모든 뉴런을 사용하여 학습하는 것이 아니라 일부 뉴런을 임의로 생략한 채로 학습을 하게 된다. 이 방법을 사용하게 되면 과적합(overfitting)을 방지할 수 있고, 특정 뉴런의 바이어스(bias)나 가중치(weight)가 큰 값을 갖게 되면 다른 뉴런들의 학습 속도가 느려지거나 학습이 제대로 안 되는 현상을 방지할 수 있게 된다.

IV. 실험 결과

본 논문에서 사용한 표지판 데이터베이스는 총 9,722장으로, 학습용 영상 총 8,561장, 실험용 영상 총 1,161장으로 구성되어 있고 총 클래스 수는 6개이다. 다양한 장소, 조명, 위치에서 촬영한 1280 × 960 크기의 영상과 인터넷 웹사이트에서 취득한 영상을 사용하였는데, 학습용 영상은 촬영한 영상과 취득한 영상에서 영상 크로핑하여 사용하였고 실험용 영상은 촬영한 영상을 그대로 사용하였다. 학습용 영상에서 각 클래스별 데이터의 개수는 남자 화장실 1581개, 비상구 2384개, 지하철 993개, 장애인 1176개, 계단 1326개, 배경 1101개이고, 실험용 영상 1,161장은 다양한 클래스의 표지판이 섞여 있으며 실험용 영상에 존재하는 모든 긍정 표본의 수는 1,318개이다. 표지판 데이터베이스의 6개 클래스에 대한 예시 영상은 그림 5와 같다. 배경 클래스의 영상은 촬영한 영상의 표지판을 제외한 영역에서 임의의 위치, 임의의 크기의 정사각형 영역을 잘라내어 생성하였다.

실험에 사용한 하드웨어는 프로세서 i7-4770K (3.5GHz), 메모리 8G RAM (DDR3), 그래픽 처리 장치 GTX 650 이고 개발 환경은 운영체제 Window 10 pro 64bit, 프로그래밍 도구 Visual Studio 2013이며 사용한 딥러닝 라이브러리는 Caffe^[10]이다.

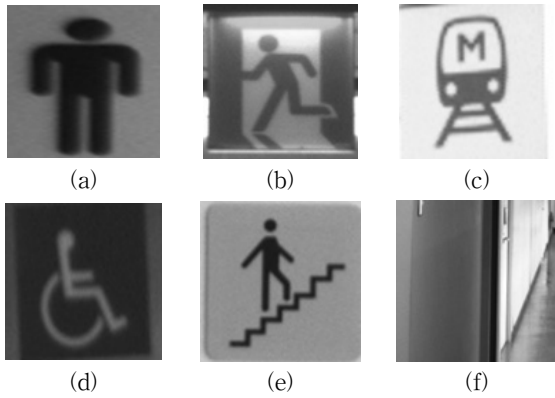


그림 5. 표지판 데이터베이스 (a) 남자 화장실, (b) 비상구, (c) 지하철, (d) 장애인, (e) 계단, (f) 배경
 Fig. 5. The sign database (a) Men's restroom, (b) Exit, (c) Metro, (d) Disabled, (e) Stair, (f) Background.

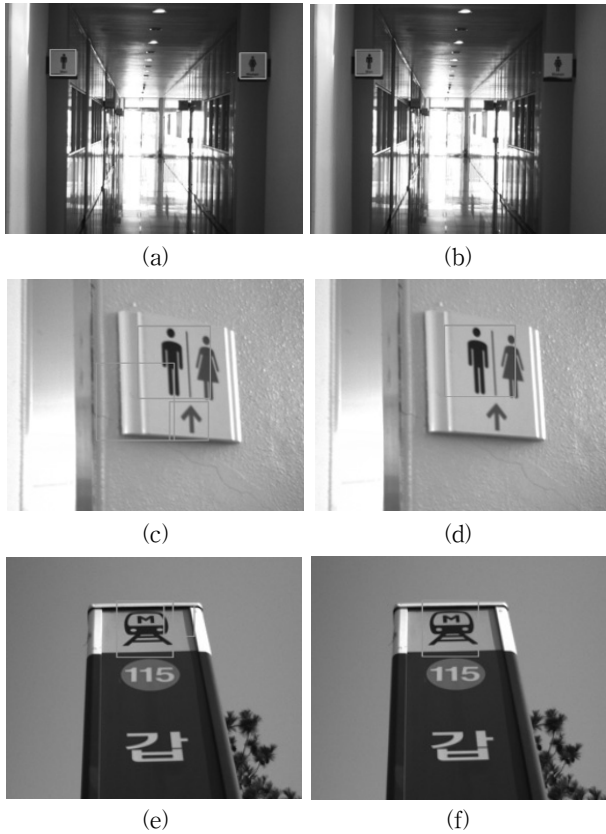


그림 6. (a)(c)(e) 표지판 검출 결과 영상, (b)(d)(f) 표지판 검출 후 합성곱 신경망을 이용해 인식한 결과 영상
 Fig. 6. (a)(c)(e) The result images of sign detection, (b)(d)(f) The result images of sign detection followed by sign recognition using CNN.

그림 6은 합성곱 신경망으로 표지판 인식하는 과정 없이 표지판 검출만 수행한 결과 영상과 표지판 검출 후 합성곱 신경망으로 표지판 인식을 수행했을 때의 결과 영상을 나타낸다. 그림 6-(a)와 그림 6-(b)를 비교해

보면 남자 화장실과 유사한 여자 화장실 표지판이 표지판 검출 단계에서 같이 검출되었지만 표지판 인식 단계에서 여자 화장실 표지판은 남자 화장실 표지판과 확연히 다른 부분이 있기 때문에 제거된 것을 알 수 있다. 그림 6-(c)와 그림 6-(d)를 비교해보면 표지판 검출 단계에서 표지판의 테두리 부분과 화살표 부분의 세로 방향 윤곽선이 많다는 점이 남자 화장실의 표지판의 세로 방향 윤곽선이 많다는 점과 유사하여 양성 오류가 발생한 모습이지만 표지판 인식을 거치면서 남자 화장실과 형태가 많이 다른 양성 오류가 제거된 것을 확인할 수 있다. 그림 6-(e)와 그림 6-(f)를 비교해보면 표지판 검출 단계에서 각도 문제 때문에 지하철 표지판의 상하 길이가 줄어들면서 지하철 마크 주변에 노란색 색상 영역이 여분으로 생겨 이 노란색 색상 영역을 지하철 표지판으로 양성 오류가 발생한 모습이지만 마찬가지로 표지판 인식을 거치면서 지하철 표지판과 형태가 많이 다르기 때문에 제거된 것을 볼 수 있다.

표지판 검출 알고리즘에서 특징으로 윤곽선 정보와 색상 정보를 사용하였기 때문에 표지판 검출 단계에서 윤곽선 정보와 색상 정보와 유사할 경우 양성 오류가 많이 발생하였다. 클래스별로 양성 오류가 많이 발생한 경우를 살펴보면, 남자 화장실 표지판과 같이 세로 방향 윤곽선이 많을 경우 세로 방향 윤곽선이 많은 영역에서 양성 오류가 많이 발생하였고, 비상구와 지하철, 장애인 표지판과 같이 특정 색상이 대부분을 차지하는 경우 색상이 유사한 영역에서 양성 오류가 많이 발생하였으며, 계단 표지판의 경우 형태가 다양함으로 인해 발생하는 양성 오류가 많았다. 표지판 검출 알고리즘을 통과한 관심 영역들에 중에 이러한 원인 때문에 양성 오류가 된 영역이 다수 있었지만 높은 인식률을 보여주는 합성곱 신경망 알고리즘을 적용하였을 때 양성 오류가 된 관심 영역들이 대부분 제거되는 것을 확인할 수 있었다. 합성곱 신경망 알고리즘을 적용하고 난 후에도 양성 오류가 남은 경우는 대부분 남자 화장실 표지판으로 인식이 된 경우였는데 보통 기둥이나 벽 같은 형태의 관심 영역이 화장실 표지판으로 잘못 인식되는 경우가 많았다. 그리고 미검출된 긍정 표본의 경우는 대부분 표지판 검출 단계에서 검출하지 못한 경우가 많았으며 크게 세 가지 형태로 분류할 수 있었다. 첫 번째로 학습 데이터에 표본이 없는 새로운 실험 데이터 표본의 경우 제대로 검출이 되지 않았고, 두 번째로 조명이 너무 어둡거나 밝은 경우 색상 차이가 커서 제대로 검출되지 않았으며, 세 번째로 비최대치 억제 기법을 거치

면서 배경이 복잡한 영상의 배경에 잡힌 강한 긍정 오류에 의해 오히려 근처의 긍정 표본이 제거되는 경우가 있었다.

제안된 방법의 성능 평가를 위해 식 (1)과 식 (2)를 사용하여 영상 1장당 양성 오류 개수(False Positives Per Image)와 미검출률(miss rate)을 측정하였다. 영상

$$\text{영상 1장당 양성 오류 개수} = \frac{\text{전체 양성 오류 개수}}{\text{전체 실험 영상 장수}} \quad (1)$$

$$\text{미검출률} = \frac{\text{미검출된 긍정 표본 개수}}{\text{전체 긍정 표본 개수}} \quad (2)$$

1장당 양성 오류 개수는 검출기가 얼마나 표지판이 아닌 물체를 표지판으로 판단하는지를 나타내는 척도이며 미검출률은 검출기가 얼마나 실제 표지판인 물체를 표지판이 아니라고 판단하는지를 나타내는 척도이다.

그림 7은 네 가지 실험에 대해서 영상 1장당 양성 오류 개수 대 미검출률 그래프를 로그 스케일로 나타낸 것이다. 네 가지 실험은 각각 표지판 검출만 수행한 실험, 표지판 검출 후에 기울기 히스토그램과 서포트 벡터 머신(HOG+SVM)을 이용한 인식을 수행한 실험, 표지판 검출 후에 합성곱 신경망을 이용한 인식을 수행한 실험, R-CNN(Regions with CNN features)^[11] 알고리즘으로 검출 및 인식을 수행한 실험이다. R-CNN 알고리즘은 영상에서 물체가 있을 법한 영역들을 영역 추출 기법을 이용하여 추출한 뒤, 추출한 영역들에 대해서만 합성곱 신경망을 적용하여 해당 영역이 어떤 물체를 나타내는지 분류하는 알고리즘이다.

실험 결과를 보면 알 수 있듯이 검출과 인식을 동시에 했을 때 검출만 했을 때보다 더 성능이 향상되었다. 검출만 했을 때보다 더 성능이 향상된 이유는 검출 단계에 배경에서 검출된 양성 오류들이 인식 단계에서 배경 클래스로 인식되거나 배경 클래스가 아니더라도 낮은 신뢰도 값을 출력하면 제거될 수 있기 때문이다. 또한 검출 단계에서 표지판으로 잘 검출되었지만 분류가 잘못됨으로 인해 발생한 양성 오류들이 인식 단계를 거치면서 제대로 분류됨으로 인해 제거될 수도 있다. 즉, 한 번 더 검증하는 과정이 있기 때문에 양성 오류 제거 확률이 높아져 성능이 향상되게 된다. 한편, 인식 단계에서 사용하는 알고리즘에 따라서도 성능이 변화하게 되는데 인식 알고리즘으로 합성곱 신경망을 사용했을 경우가 기울기 히스토그램과 서포트 벡터 머신을 사용했을 경우보다 더 성능이 좋다. 합성곱 신경망 알고리

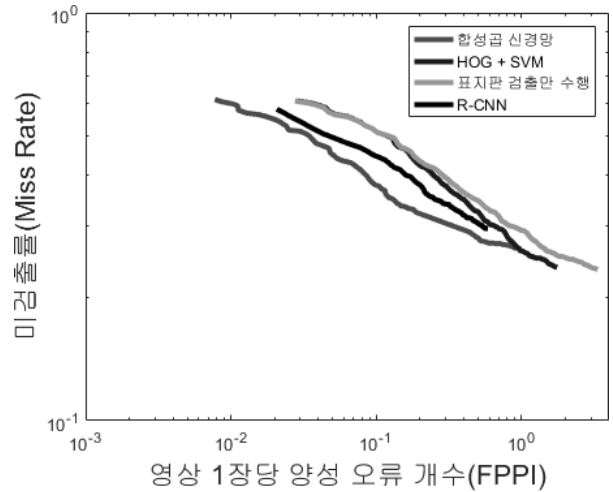


그림 7. 영상 1장당 양성 오류 개수 대 미검출률 그래프
Fig. 7. The graph of False Positive Per Image vs Miss Rate.

즘의 모델이 더 복잡한 구조를 가지고 있기 때문에 표지판의 세밀한 특징을 더 잘 잡아내 좋은 성능을 보이는 것으로 분석된다. 마찬가지로 검출 단계에서 사용하는 알고리즘에 따라서도 성능이 변화하는데 아다부스트 분류기를 사용했을 경우가 영역 추출 기법을 사용했을 경우보다 더 성능이 좋다. 아다부스트 분류기의 경우 영상의 크기를 변화시켜가며 슬라이딩 윈도우 기법을 적용하여 모든 영역을 훑으면서 표지판의 존재 여부를 검토하지만 영역 추출 기법의 경우 영상의 전체적인 특징을 이용하여 표지판이 있을 법한 영역을 대략적으로 추정하기 때문에 실제로 표지판이 있는 부분을 놓칠 확률이 더 높은 것으로 분석된다.

V. 결 론

본 논문에서는 시각장애인의 보행을 보조하기 위해 영상에서 시각장애인에게 필요한 주요 시설의 위치를 나타내는 표지판을 검출 및 인식하는 시스템을 제안하였다. 제안된 방법에서는 미검출률과 양성 오류 개수를 감소시키기 위해 시스템을 표지판 검출 단계와 표지판 인식 단계로 나누어서 설계하였다. 표지판 검출 단계에서는 미검출된 표지판의 개수가 최대한 감소하도록 아다부스트 분류기를 설계하였고 표지판 인식 단계에서는 딥러닝 알고리즘 중 하나인 합성곱 신경망을 사용하여 인식 정확도를 높이고 양성 오류를 제거하고자 하였다.

실험 결과, 표지판 검출 단계에서 발생한 양성 오류들이 표지판 인식 단계를 통해 한 번 더 검증받게 되고 이를 통해 제거가 가능해지면서 양성 오류 개수의

감소로 인한 성능 향상을 확인할 수 있었다. 그리고 복잡한 구조의 합성곱 신경망을 표지판 인식 알고리즘으로 사용함으로써 표지판의 특징을 잘 잡아낼 수 있었고, 간단한 구조의 인식 알고리즘을 사용했을 때보다 성능이 더 향상된 것을 확인할 수 있었다. 또한, 슬라이딩 윈도우 기법을 적용해 모든 영역을 훑으며 표지판을 검출함으로써 대략적인 영역 추출 기법보다 표지판의 위치를 더 잘 찾아내는 것을 확인할 수 있었다.

하지만 조명 변화 때문에 발생한 역광, 반사, 그림자 등에 의해 표지판의 색상이 학습 영상들과 다르게 나타날 경우 검출률이 떨어지고, 학습 영상에 포함되지 않은 변형된 표지판의 경우도 검출률이 떨어지는 것을 확인할 수 있었다. 따라서 향후 연구로 조명 변화에 대한 시스템의 민감도를 최소화 할 수 있는 연구와 표지판의 변형에 강인한 시스템에 대한 연구가 필요하다.

REFERENCES

- [1] C. Yoon, S. Jang and M. Park, "Real-time road sign detection using vertical plane and adaboost," Journal of the Institute of Electronics Engineers of Korea SC, vol. 46, no.5, pp. 29-37, Sep. 2009.
- [2] J. Kim and J. Park, "Traffic sign detection using the HSI eigen-color model and invariant moments," Journal of the Institute of Electronics Engineers of Korea CI, vol. 47, no.1, pp. 41-51, Jan. 2010.
- [3] B. Tian, R. Chen, Y. Yao and N. Li, "Robust traffic sign detection in complex road environments," Vehicular Electronics and Safety (ICVES), 2016 IEEE International Conference on. IEEE, pp. 1-5, Jul. 2016.
- [4] Y. Bengio, A. Courville and P. Vincent, "Representation learning: A review and new perspectives," IEEE transactions on pattern analysis and machine intelligence, vol.35, no.8, pp. 1798-1828, 2013.
- [5] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li and S. Hu, "Traffic-Sign Detection and Classification in the Wild," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2110-2118, 2016.
- [6] M. Peemen, B. Mesman and H. Corporaal, "Speed sign detection and recognition by convolutional neural networks," 8th International Automotive Congress, pp. 162-170, 2011.
- [7] P. Dollár, R. Appel, S. Belongie and P. Perona, "Fast feature pyramids for object detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.36, no.8, pp. 1532-1545, 2014.
- [8] P. Viola and M. Jones, "Robust Real Time Object Detection," IEEE ICCV Workshop Statistical and Computational Theories of Vision, Jul. 2001.
- [9] D. CireşAn, U. Meier, J. Masci and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," Neural Networks, vol.32, pp. 333-338, 2012.
- [10] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," Proceedings of the 22nd ACM international conference on Multimedia. ACM, pp. 675-678, Nov. 2014.
- [11] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587, 2014.

저 자 소 개



전 태 재 (학생회원)

2014년 연세대학교 전기전자공학과 학사 졸업.

2014년~현재 연세대학교 전기전자공학과 석박사통합과정.

<주관심분야: 심층 인공 신경망, 컴퓨터비전>



이 상 윤 (평생회원)

1987년 연세대학교 전기전자공학과 학사 졸업.

1989년 연세대학교 전기전자공학과 석사 졸업.

1999년 Georgia Tech. 전기 및 컴퓨터공학과 박사 졸업.

1989년~2004년 KT 선임연구원.

2004년~현재 연세대학교 전기전자공학과 교수.

<주관심분야: 컴퓨터비전, 영상인식, 영상부호화>