

정규논문 (Regular Paper)

방송공학회논문지 제20권 제6호, 2015년 11월 (JBE Vol. 20, No. 6, November 2015)

<http://dx.doi.org/10.5909/JBE.2015.20.6.938>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

랜덤 포레스트 분류기 기반의 컨벌루션 뉴럴 네트워크를 이용한 속도 제한 표지판 인식

이 은 주^{a)}, 남 재 열^{a)}, 고 병 철^{a)†}

Speed-limit Sign Recognition Using Convolutional Neural Network Based on Random Forest

EunJu Lee^{a)}, Jae-Yeal Nam^{a)}, and ByoungChul Ko^{a)†}

요 약

본 논문에서는 외부압력에 의한 외형 손상이나 빛의 방향에 따른 색상 대비변화 등에 견고한 영상기반 속도 제한 표지판 인식 시스템 설계를 제안한다. 속도 제한 표지판 인식을 위해서 최근 패턴 인식 분야에서 뛰어난 성능을 보여주고 있는 CNN (Convolutional neural network)을 사용한다. 하지만 기존의 CNN은 특징 추출을 위해 다수의 은닉층이 사용되고 추출된 결과에 대해 MLP(Multi-layer perceptron) 등과의 완전 연결(fully-connected) 방식을 사용함으로 학습과 테스트 시간이 많이 걸리는 단점이 있다. 본 논문에서는 이러한 단점을 줄이기 위해 2계층의 CNN을 구성하고 패턴 분류를 위해 랜덤 포레스트(Random forest)를 결합하여 완전 연결이 아닌 랜덤 연결 방식을 적용하였다. GTSRB(German Traffic Sign Recognition Benchmark)데이터의 교통안전표지판 중에서 8개 속도 제한 표지판 데이터를 사용하여 제안하는 방식이 SVM (Support Vector Machine)이나 MLP 분류기를 적용할 때 보다 성능이 우수함을 입증하였다.

Abstract

In this paper, we propose a robust speed-limit sign recognition system which is durable to any sign changes caused by exterior damage or color contrast due to light direction. For recognition of speed-limit sign, we apply CNN which is showing an outstanding performance in pattern recognition field. However, original CNN uses multiple hidden layers to extract features and uses fully-connected method with MLP(Multi-layer perceptron) on the result. Therefore, the major demerit of conventional CNN is to require a long time for training and testing. In this paper, we apply randomly-connected classifier instead of fully-connected classifier by combining random forest with output of 2 layers of CNN. We prove that the recognition results of CNN with random forest show best performance than recognition results of CNN with SVM (Support Vector Machine) or MLP classifier when we use eight speed-limit signs of GTSRB (German Traffic Sign Recognition Benchmark).

Keyword : Convolutional Neural Network, Random forest, speed-limit sign recognition , feature extraction, ADAS

a) 계명대학교 컴퓨터공학과(Dept. of Computer Engineering, Keimyung University)

† Corresponding Author : 고병철(ByoungChul Ko)

E-mail: niceko@kmu.ac.kr

Tel: +82-53-580-6275

ORCID: <http://orcid.org/0000-0002-7284-0768>

※ 본 연구는 산업통상자원부·한국산업기술진흥원 지정 계명대학교 전자화자동차부품지역혁신센터(B0008866)의 지원에 의한 것입니다.

· Manuscript received September 12, 2015; revised November 4, 2015; accepted November 4, 2015.

1. 서론

지능형 자동차 분야에서 운전자 지원 시스템(Advanced Driver Assistant System, ADAS)은 운전자와 보행자를 사고로부터 보호 할 수 있는 핵심적인 기술이다^[1]. 특히 카메라를 이용한 컴퓨터 비전 기술은 보행자 검출, 차선 인식, 표지판 인식, 줄임 운전자 인식 등의 다양한 ADAS 분야에 적용되어 오고 있다. ADAS 관련 기술 중 특히 교통 표지판 인식은 부주의한 운전자에게 ‘방향 전환’, ‘속도 제한’ 등의 정보를 인지 시켜줌으로써 교통사고를 예방하고 무인 자율주행 시스템을 보조할 수 있는 기능을 제공한다.

EU^[2]의 결과에 따르면 충돌 10미터 전에 자동차 주행 속도를 시속 50km에서 31km로 줄이면 사망률을 33%에서 5%로 줄일 수 있는 것으로 연구 되었다. 따라서 교통 표지판 인식 방법 중 속도 표지판 인식은 보행자의 사망률을 줄일 수 있다는 측면에서 매우 중요한 연구 분야이다. 일반적으로 도로의 표지판은 내비게이션 시스템에서 자동차의 주행 속도와 GPS정보를 결합하여 예측해 주지만 내비게이션이 실시간으로 업데이트 되지 않는다면 도로가 추가 되었거나 제한속도가 바뀌었을 때 운전자에게 정확한 정보를 주지 못하는 문제점이 있다. 이러한 문제점을 해결하기 위해 컴퓨터 비전 기술을 이용한 속도 제한 표지판을 인식하는 연구가 꾸준히 진행되어 왔다.

속도 제한 표지판 인식의 주요 특징벡터로는 주로 HOG(Histogram of Gradient)^[3]와 RSD(Radial Symmetry Detection)^[4]방법이 주로 사용되며, 인식에서는 SVM^[5]과 인공신경망(Neural Networks)^[6], 랜덤 포레스트^[7]방법이 사용되어 왔다.

Mathias^[8]등은 검출된 관심 영역에서 밝기의 분포 방향 기반의 HOG 히스토그램을 얻어 이를 특징 벡터로 추출하고 추출한 문자의 분류를 위해 Multi-Class SVM을 사용하여 학습시킨 데이터와 비교해 해당 문자가 어느 속도 클래스에 속해 있는지 분류하는 방법을 사용하였다. 하지만 HOG 방법은 우수한 성능에 비해 큰 특징 차원 때문에 학습과 테스트 시간이 많이 걸리므로 실시간 처리에 적합하지 않다.

Barns^[9]등은 원형 허프 변환에 기반 한 RSD(Radial Symmetry Detector)방법을 이용하여 표지판의 중심과 지

름을 구하고, 정규화된 교차 상관관계(Cross Correlation)를 이용하여 속도 제한 표지판을 분류하는 알고리즘을 제안하였다. 하지만, 이 방법은 관심 영역이 교통 표지판의 크기와 일치해야 하고, 정면으로 촬영된 영상이 아니 라면 형태 왜곡이 발생하여 인식이 떨어지는 문제점이 있다.

Aoyagi와 Asakura^[10]은 운전자 알고리즘과 제안된 필터를 이용하여 속도 표지판의 위치를 추정하고 후보 표지판을 인공 신경망에 적용하여 최종 표지판을 속도별로 분류하는 알고리즘을 제안하였다. 인공신경망을 이용하는 방법은 영상에 대해 전처리 없이 영상 자체를 특징으로 사용할 수 있다는 장점이 있지만, 이 경우 특징의 개수가 너무 많아 연산량이 많다는 단점이 있다.

패턴 인식 분야에서는 1990년대 후반부터 영상과 같이 대량의 데이터를 의미 있는 데이터의 형태로 표현하고 학습하고자 하는 노력으로 CNN과 같은 딥 러닝(Deep Learning) 기법들이 다양한 영상 인식 분야에 적용되기 시작했으며, 딥 러닝의 효용성을 입증하기 위한 많은 연구가 진행되어 왔다. 대표적으로 CNN을 기반으로 만들어진 Caffe 라이브러리^[11]는 필기체 숫자인 성능 테스트용 데이터인 MNIST(Mixed National Institute of Standards and Technology database)^[12]를 학습시키고 결과를 테스트 해 볼 있도록 설계되었다. MNIST는 사람이 손으로 쓴 0~9까지 숫자들의 모음으로 딥 러닝에서 아주 흔하게 사용되는 데이터 셋 중 하나이다.

CNN을 이용한 인식 연구는 컴퓨터 비전, 음성 인식, 영상 인식 등의 분야에 적용되어 성공적인 인식 성능을 보여 주고 있으며 최근에는 각 응용분야에 맞게 다양한 형태로 변형 되어 적용되어 오고 있다.

딥 러닝기술을 기반으로 하는 CNN은 다층 인공 신경망의 구조를 나타내고 있기 때문에 사용자의 간섭이 적고 얼굴 인식과 같은 일부 연구 분야에서는 인간의 인식능력과 유사한 결과를 보여 것으로 알려져 있지만, 계층을 학습 시킬 때 계층의 구성이나 필터의 종류, 초기 가중치 등 많은 매개변수(parameters)에 따라 성능 차이가 나게 되고 계산에 필요한 매개변수의 양에 따라 학습과 테스트 시간 또한 많은 차이를 보이는 문제점도 동시에 지니고 있다.

본 연구에서는 복잡한 CNN의 계층 구조를 2계층으로 축

소하여 CNN의 다층 구조로 인한 문제점을 해결하고 각 계층마다 의미 있는 4개의 필터만을 사용하여 특징을 축소하도록 하였다. 본 논문에서 제안하는 구체적인 알고리즘은 다음과 같다.

- 입력 속도 표지판에 포함된 다양한 조명을 전처리 단계를 통해 개선
- CNN의 각 계층에서 사용되는 컨벌루션 필터를 결정하기 위해 속도 표지판의 특성을 잘 나타낼 수 있도록 필터뱅크 업데이트를 반복 수행하여 속도 제한 표지판 인식에 최적의 필터를 구성
- 추출된 특징 벡터를 완전 연결 방식의 MLP대신에 랜덤 연결 방식의 랜덤 포레스트와 결합하여 처리 속도를 향상 시키고 성능을 높이도록 알고리즘을 설계

본 논문의 구성은 다음과 같다. II에서는 속도 제한 표지판 인식을 위한 2계층으로 구성된 CNN의 구조에 대해서 소개하고 III에서는 제안된 방법의 효율성에 대한 실험 및 성능 평가 방법을 보인다. 마지막으로 IV에서는 결론 및 향후 연구 방향에 대하여 기술한다.

II. 속도 제한 표지판 인식을 위한 빠른 2계층 CNN

1. 기존 CNN의 구성

전통적인 CNN방법은 기존의 인공신경망을 연속적으로

나열하는 MNN (Multi-layer Neural Network) 방법을 이용하고 있다. MNN에서의 영상인식은 영상을 구성하는 픽셀 자체를 입력으로 사용하는 경우가 많다. 하지만 입력된 영상의 모든 픽셀을 출력까지 완전 연결(fully-connected)로 구성할 경우는 다음과 같은 문제점을 가질 수 있다.

- 1) 데이터 차원과 매개변수 문제: MNN에서 성능 향상을 위해 은닉층(hidden layer)의 수를 결정하게 되는데 은닉층의 수가 늘어나면 영상 픽셀 수의 거듭제곱으로 연결이 늘어나고 훈련 시켜야 하는 매개변수 수도 늘어나게 된다.
- 2) 영상 왜곡 문제: 주변 환경에 의해 발생된 영상 왜곡이 그대로 입력에 반영된다.
- 3) 영상 특성의 반영 문제: 영상은 2차원 정보로 주변 픽셀과 서로 높은 연관성을 가지지만 단순한 벡터 형식의 연결 방법은 유용한 공간 정보를 반영하지 못한다.

LeCun 등이 제안한 CNN^[13]는 역전파 알고리즘을 적용하여 학습 기능을 추가한 다-계층구조로 된 신경망이며 1998년 필기체 인식을 위해 처음 제안 되었다. 그림 1은 [13]에 의해 제안된 전통적인 CNN 구조이다.

CNN은 특히 ILSVRC (Imagenet Large Scale Visual Recognition Challenge)에서 물체 인식에 대단히 뛰어난 성능이 검증되면서 다양한 분야로 확장 중이다^[9]. 전통적인 CNN의 각 계층은 지역적 수용 필드(Local receptive field), 공유 가중치(Shared weight), 서브샘플링(Subsampling, Pooling)의 세 가지 아이디어가 적용 되었으며 지역적 수용 필드, 공유 가중치를 반영하는 컨벌루션 계층(Convolution

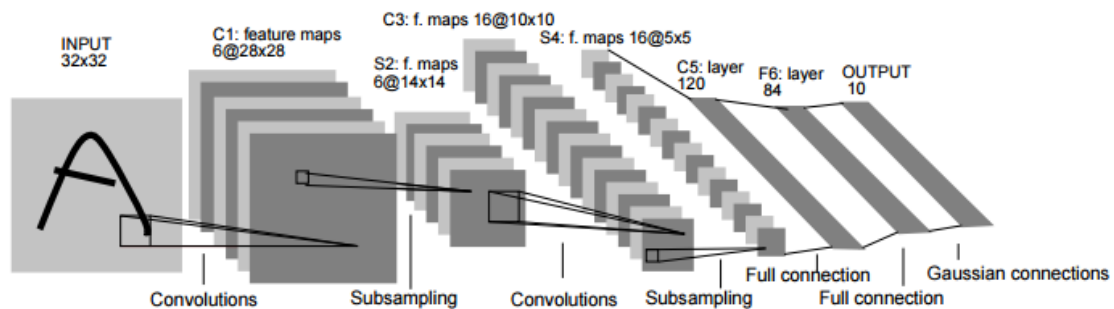


그림 1. 필기체 인식을 위해 설계된 전통적인 CNN 구조^[13]
 Fig. 1. Traditional CNN structure designed for handwriting recognition

layer)과 서브샘플링 계층이 번갈아 수행되며 완전 연결 계층 (Fully Connected Layer)을 통해 분류 수행하는 계층 모델이다. 기본적인 이 두 가지의 층은 여러 층으로 적재가 가능하다. 다음은 각 계층에 대한 설명이다.

1.1 컨벌루션 계층 (Convolution layer)

컨벌루션 계층은 MNN과는 달리 인접한 픽셀끼리의 지역 (Local)특성을 반영하고자 2차원 배열 형태의 입력 이미지에 $N \times N$ 크기의 필터를 컨벌루션 연산한다. 이러한 연산은 객체의 위치나 크기에 영향을 받지 않고 입력된 영상에서 찾아내 고자 하는 객체의 에지나 코너, 선의 끝과 같은 특징을 특징 맵 (Feature map)에 표현하기 위한 방법이다. 컨벌루션 되는 필터의 개수가 많아지면 다양한 종류의 특징을 추출 할 수도 있다. 컨벌루션 연산은 그림 2와 같은 방법으로 영상내의 모든 픽셀에 대해 반복적으로 처리한다^[14].

2차원 입력 영상에 대해서 $N \times N$ 크기의 필터를 모든 가능한 위치에서 컨벌루션 연산이 수행될 수 있도록 이동시키므로 출력되는 특징 맵의 크기는 (입력 영상의 크기 - 필터 크기 +1) 이 된다. 따라서 32×32 입력 영상에 대해 3×3 크기의 필터를 적용할 경우 컨벌루션 후의 영상 크기는 30×30 픽셀이 된다.

1.2 서브샘플링 계층 (Subsampling, Pooling layer)

영상의 경우 픽셀개수가 많으므로 크기를 줄이기 위해 영상을 축소하는 과정을 거치는데 이것을 서브샘플링 (Subsampling, Pooling)이라고 부른다. 일반적으로 입력 픽셀을 2×2 크기의 부 영역으로 분할하고 각 서브 영역에서 픽셀의 최댓값 또는 평균값을 추출하여 2D 영상의 한 점으로 매핑 한다. 서브샘플링을 거치면 국소적인 영상의 변형에 강인한 특징 벡터를 추출할 수 있으며, 2×2 의 픽셀들이 하나의 픽셀로 줄어들기 때문에 이론적으로 상위단계의 추상화 효과가 생기게 된다. 일반적인 서브샘플링 방법은 주변의 픽셀들의 값을 비교하여 주로 가장 큰 값을 취하는 Max-pooling 방법을 사용한다.

1.3 완전 연결 계층 (Fully connected (FC) layer)

마지막으로 적용되는 단계이며 서브샘플링 계층에서 나온 특징들을 이용해서 영상을 클래스별로 분류를 할 때 사용되는데, 일반적인 인공지능망의 입력처럼 동작한다. 영상 인식을 위한 분류기로는 상관계수 (Correlation) 기반 템플릿 매칭을 이용한 방법^[5], 신경망을 이용한 방법^[14], SVM을 이용한 방법^[5], 선형판별법을 이용한 방법^[16]이 있으며 이외에도 오류정정 출력코드(Error Correcting Output Code)를 이용한 방법^[17]이 있다. 본 논문에서는 추출된 특징을 기반으로 결정 트리(Decision tree)를 구성하여 분류하는 랜덤 포레스트 방법을 제안한다.

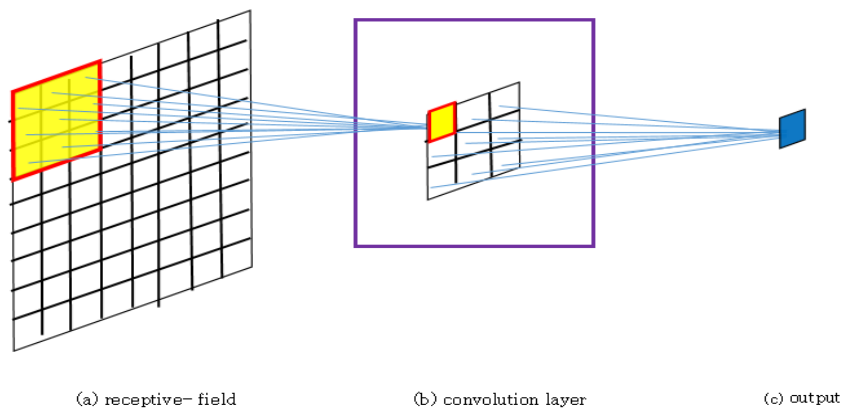


그림 2. 필터를 이용한 컨벌루션 과정, (a) 수용필드 (b) 컨벌루션 계층 (c) 출력 값
 Fig. 2. Convolution Operation using a filter, (a) Receptive field, (b) Convolution layer, (c) Output

2. 제안하는 랜덤 연결 방식의 2계층 CNN

2.1 CNN 계층 구성

본 연구는 속도 제한 표지판을 인식하기 위하여 그림 3과 같은 2계층의 CNN을 구성하였다. 제안하는 CNN 계층은 입력 계층, 1차 컨벌루션 계층 (C1), 1차 서브샘플링 계층 (S1), 2차 컨벌루션 계층 (C2), 2차 서브샘플링 계층 (S2), 랜덤 포레스트 계층으로 구성 되어 있다. 일반적으로 컨벌루션 계층과 서브샘플링 계층을 1개의 통합 계층으로 정의됨으로 제안된 CNN은 총 2개의 계층으로 구성되어 있다.

2.1.1 입력 계층

속도 표지판 영상은 카메라와의 거리에 따라 다양한 크기를 가지므로 검출된 속도 표지판 영역을 32×32 로 정규화 시키고 낮은 명암 대비를 보이는 영상들은 화질을 향상시키기 위하여 히스토그램이 모든 범위의 화소 값을 포함하도록 상위 1% 하위 1%를 제외하는 히스토그램 스트레칭 (Histogram stretching)을 수행한다. 그림 4는 낮은 명암 대비를 보이는 영상에 대한 히스토그램 스트레칭 적용 전과 적용 후의 결과를 나타낸다. 전처리로 보정된 영상은 C1 계층의 특징 맵 생성을 위한 데이터로 입력된다.



그림 4. (a) 히스토그램 스트레칭 적용 전 과 (b) 적용 후의 결과
Fig. 4. (a) original speed-limited signs and (b) histogram stretched signs

2.1.2 C1 계층

컨벌루션 계층의 첫 번째 계층인 C1 계층은 4개의 특징 맵으로 구성되어 있다. C1 계층에서는 3×3 크기의 필터를 적용하여 특징 맵을 생성한다. 컨벌루션 연산 시 경계를 고려하여 각 특징 맵은 30×30 크기를 갖는다. 본 연구에서 적용한 3×3 필터는 그림 5와 같다. 필터는 9개의 값으로 이루어져 있으며 9개의 랜덤 값으로부터 최적의 필터 계수를 추정하는 방법은 LeCun 등이 제시한 방법^[14]과 동일한 방법을 사용하였다. 먼저 0~255의 범위를 9개의 영역으로

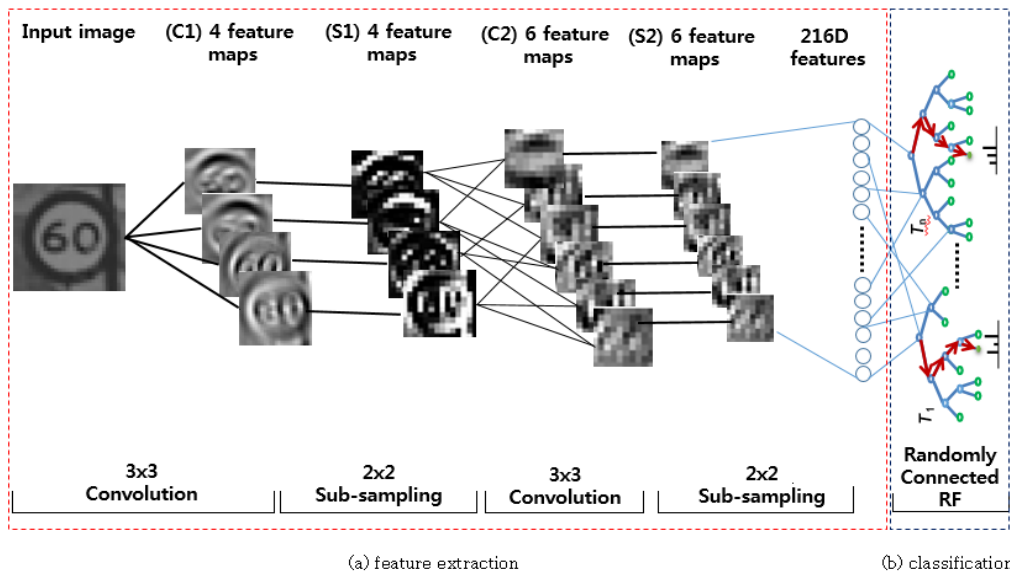


그림 3. 제안하는 2계층 CNN 구조 (a) 특징 추출을 위한 컨벌루션 계층과 서브샘플링 계층 (b) 분류를 위한 랜덤 포레스트 계층
Fig. 3. The proposed 2-layer CNN architecture, (a) convolution and subsampling layer for feature extraction, (b) random forest layer for classification

나누고 이 9개의 영역에서 각각 임의의 값을 추출해 낸다. 그 값의 음수와 양수 값을 이용하여 하나의 필터 값을 구성하고, 같은 방법으로 300번의 필터 생성 과정을 거쳐 만들어진 필터들 중에서 성능이 우수한 필터를 선택한다.

그림 5는 랜덤하게 생성된 300개의 필터에서 실험을 통해 정확도가 80% 이상인 필터 중에서 3개의 필터를 나열하였다. 본 논문에서는 최상위 필터 중에서 속도 표지판의 특성을 잘 나타내고 가장 좋은 성능을 보여준 그림 5-(c)의 4개 필터를 1차와 2차 계층에서 동일하게 사용하였다.

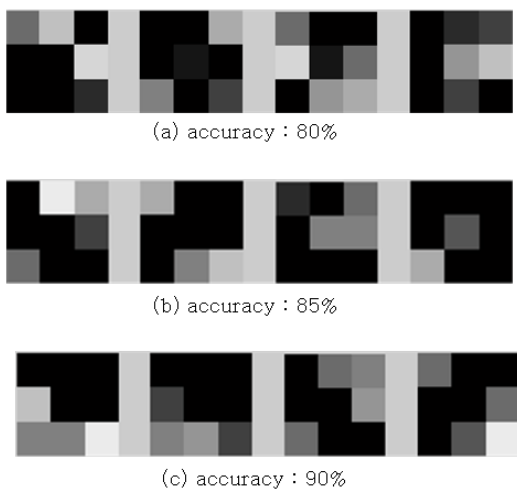


그림 5. C1 계층 필터
 Fig. 5. C1 layer filters

학습 단계에서 그림 5-(c)로 구해진 컨벌루션 필터는 입력 영상에 적용되어 4개의 특징 맵을 생성한다. 컨벌루션 연산 시에 필터와 입력 영상은 그림 3의 C1 단계에서 보듯이 1:1 매핑으로 특징 맵을 구성한다. 컨벌루션 필터링 된 값들은 식 (1)과 같이 비선형 변환(Non-linear activation)함수인 ReLU(Rectified Linear hidden Unit)^[18]를 이용하여 활성화 된다. 컨벌루션 필터링 된 값은 입력 값(x)으로 사용되며 $x > 0$ 일 때는 $f(x) = x$, $x \leq 0$ 일 때는 $f(x) = 0$ 으로서 ReLU는 입력 값으로 어떤 양수를 넣기만 하면 학습이 일어난다. ReLU를 적용함으로써 뉴런이 선형적으로 활성화 되어 큰 값을 가질 수 있게 함으로써 경사도가 상수가 되도록 하여 학습 과정에서 문제가 되던 Gradient vanishing 문제가 완화되었다^[19].

$$f(x) = \max(0, x) \quad (1)$$

2.1.3 S1 계층

CNN에서 컨벌루션 만큼 중요한 연산이 서브샘플링 연산이다. 첫 번째 컨벌루션 계층에서 생성된 특징 맵이 위치 변화에 따라 특징 값이 변하지 않도록 고려되어 졌다면, 특징 맵의 해상도를 줄이는 방법으로 영상의 노이즈나 왜곡도 고려되어야 한다. 본 연구에서는 C1 계층의 각 특징 맵을 2x2픽셀 크기의 부 영역으로 나누고 각 부 영역에 대해 서브샘플링 과정을 거치면서 15x15픽셀 크기로 줄어든 4

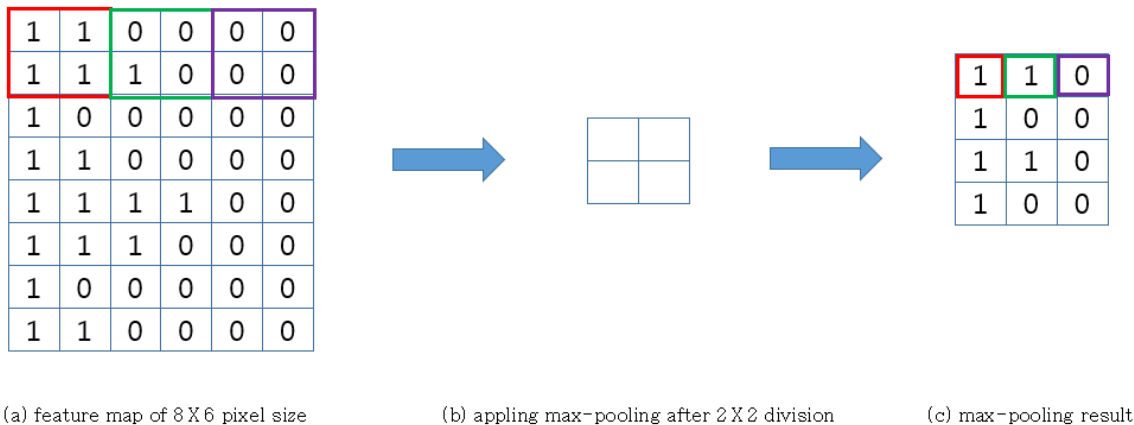


그림 6. Max pooling 방법 (a) 8x6크기의 특징 맵, (b) 2x2 부 영역으로 분할 뒤 max-pooling 적용, (c) max-pooling 적용 결과
 Fig. 6. Max pooling method (a) feature map of 8x6 pixel size, (b) applying max-pooling after 2x2 division, (c) max-pooling result

개의 특징 맵을 생성하도록 한다. 서브샘플링을 위한 Mean-pooling, Max-pooling 방법 중, 본 논문에서는 성능이 가장 좋은 것으로 알려진 Max-pooling 방법을 사용한다. 서브샘플링 계층은 영역을 서로 겹치지 않도록 배치하므로 영역의 크기에 반비례하여 특징 맵의 크기가 작아지게 된다. 그림 6은 본 논문에서 사용한 Max-pooling하는 방법을 도식화 한 것이다.

2.1.4 C2 계층

C2 계층은 컨벌루션 계층으로 C1 계층의 특징 맵 생성과는 다르게 구성된다. 우선 C1 계층에서 사용한 동일한 필터를 S1 특징 맵에 적용하고 컨벌루션 결과들의 부분적인 결합에 의해 총 6개의 특징 맵을 생성한다. 본 논문에서는 C2 계층에서의 특징 맵을 위해 S1 계층의 특징 맵들에 대해 컨벌루션 필터를 적용하고 이들을 가중 결합 방식을 적용하는데 가중치의 크기에 따라 생성되는 특징 맵의 특성이 달라짐으로 시스템의 인식 성능 향상을 위해서는 최적의

가중치를 결정해야 한다. 각 특징 맵별로 가중치를 결정하기 위해 특징 맵을 구성하기 위해 사용된 S1 계층의 특징 맵 가중치의 합이 1이 되도록 랜덤하게 구성하고 50번의 반복적 학습 후에 가장 좋은 성능을 보이는 가중치를 선택한다. 본 논문에서는 그림 7과 같이 C2 계층의 특징 맵을 생성하기 위해 2개의 컨벌루션 결과를 가중 결합하는 방법을 사용한다.

2.1.5 S2 계층

그림 3의 C2 계층과 연결된 S2 계층은 S1 계층과 마찬가지로 서브샘플링 작업을 수행한다. 따라서 C2 계층의 각 특징 맵을 2×2 픽셀 크기의 부 영역으로 나누고 max-pooling을 통해 6×6 픽셀 크기로 6개의 최종 특징 맵을 생성한다.

2.2 RF 계층 (Randomly connected RF)

기존 CNN은 분류 단계에서 MLP 신경망 구조를 분류기로 채택하고 있다. 하지만 많은 양의 데이터는 초기 매개변수에 대한 선행 학습이 이루어져야 한다는 전제가 있어야 정확한 분류가 가능하므로 이러한 신경망 구조 대신에 본 연구에서는 랜덤 포레스트를 분류기로 사용하여, 전체 특징을 사용하지 않고 랜덤 특징을 선출하여 분류에 적용하는 방식을 사용한다. 랜덤 포레스트 분류기는 이진 결정 트리들의 결합을 기본으로 하는 분류기로서 높은 차원의 특징벡터에 대해 SVM 또는 Adaboost보다 빠른 속도로 분류할 수 있다는 장점이 있다.

특징 생성 단계의 최종 계층인 S2 계층의 특징 맵을 선형적으로 연결 하면 총 216차원(특징 맵 차원: 36 × 특징 맵 수: 6)이 생성되며 이를 랜덤 포레스트 분류기의 입력으로 사용한다. 본 논문에서는 8개 클래스 총 12,550장의 학습 데이터로부터 속도 제한 표지판을 위한 특징을 생성한 후 랜덤 포레스트의 각 결정 트리 100개가 구축되도록 설계되었다. 랜덤하게 생성된 트리가 모두 우수한 결과를 나타낸 것은 아니므로 분류 성능이 상대적으로 뛰어난 N개의 트리를 선택할 필요가 있다. 상위 N개의 트리를 선택하기 위해 Adaboost와 유사한 방법으로 부스티드 랜덤 포레스트(BRF, Boosted Random Forest)^[20]를 적용한다. 본 논문

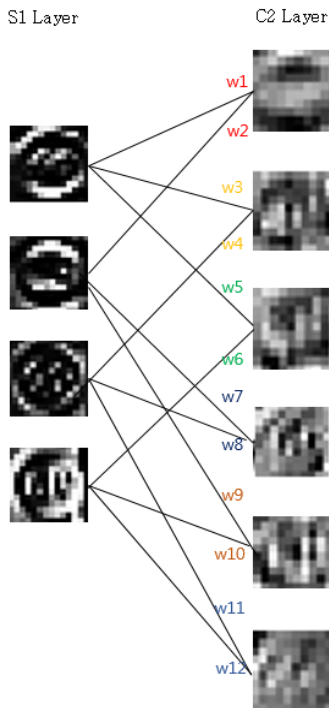


그림 7. C2 계층의 특징 맵 생성을 위한 가중치 결합 방법
Fig. 7. Weight combination method for C2 layer feature map

서는 80개의 결정 트리를 결합하여 포레스트를 구성하였다. 초기 포레스트에서 결정트리가 생성되면 분류 성능이 떨어지는 트리를 반복적으로 제거하는 방식으로 트리의 수를 최적화 하여 분류 성능을 높이고 분류 시간을 줄이도록 하였다. BRF가 구성된 이후에 새롭게 입력된 테스트 영상에 대해서 C1~S2까지의 단계를 거쳐 특징벡터를 생성하고 각 트리에 값을 입력 시켜 나온 확률 값들을 그림 8과 같이 누적하여 최종 분류 클래스를 결정하도록 한다.

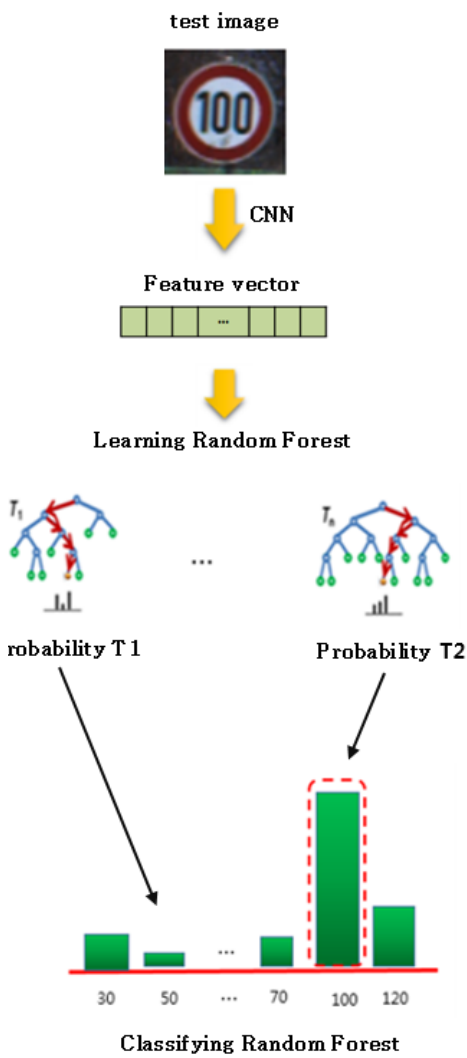


그림 8. 랜덤 포레스트를 이용한 속도 제한 표지판 분류
 Fig. 8. Speed-limit sign classification using Random forest

III. 실험 및 성능 평가

1. 실험 데이터

제안된 2계층 CNN의 성능을 측정하기 위하여 독일의 GTSRB (German Traffic Sign Recognition Benchmark)^[21] 데이터를 사용하였다. GTSRB 데이터는 교통표지판의 인식을 위해서 표지판 영역으로만 제작된 영상으로서 비디오에서 초당 30 샘플들이 추출되었다. 영상은 15×15픽셀 부터 180×180픽셀 까지 다양한 해상도로 잘려 있으며, 총 51,883개의 데이터 중에서 속도 제한 표지판은 13,490장으로 조명 변화, 영상 왜곡, 색상 변화, 회전등이 포함된 영상을 다수 포함하고 있다. GTSRB 데이터는 속도 표지판 이외에 여러 교통 표지판을 포함하고 있으므로 교통표지판 중에서 속도 제한 표지판에 해당하는 20, 30, 50, 60, 70, 80, 100, 120에 대한 총 8개의 클래스를 사용 하였다. 본

표 1. GTSRB 실험 데이터 구성
 Table 1. GTSRB Dataset Specification

GTSRB Data		Resolution of speed - limit signs	Number of signs	Image
Training set	20	25×26 ~161×168	200	
	30		2200	
	50		2250	
	60		1400	
	70		1900	
	80		1800	
	100		1400	
	120		1400	
Test set	20	29×28 ~121×130	40	
	30		150	
	50		150	
	60		120	
	70		115	
	80		150	
	100		115	
	120		100	

논문은 일반 CNN과 다르게 GPU를 사용하지 않고 CPU만으로 빠르게 동작할 수 있는 알고리즘을 개발하는 것이 목적으로 CPU만을 사용하여 테스트 하였다. 본 시스템은 Intel i7 CPU 3.6GHZ 메모리 32GB Windows 7 64비트에서 테스트 되었으며 매트랩 R2010a로 개발되었다.

표 1은 실험에 사용된 영상에 대한 설명을 보여준다. 표 1과 같이 학습 및 테스트 데이터의 영상 크기가 매우 다양함으로 CNN에 적용되기 전에 32×32로 정규화 되었다.

2. 전처리에 따른 성능 비교

표 1에서 보는 것과 같이 실험데이터는 다양한 실제 도로 환경에서 촬영 되었으므로 매우 다양한 배경과 조명 환경을 가지고 있다. 특히 조명은 인식 성능과 매우 밀접한 관련이 있으므로 다양한 조명을 가지는 영상의 밝기를 조절하기 위해 히스토그램 스트레칭 방법을 적용하였다. 표 2에서처럼 전처리 방법은 실험을 통해 결정 되었다.

표 2 전처리에 따른 속도 제한 표지판 인식 성능 비교
Table 2 Comparison of speed-limit sign recognition according to the preprocessing

Preprocessing method	original image	histogram equalization	histogram stretching
Average precision	85.73%	79.75%	91.2%

전처리 방법 중에서 히스토그램 스트레칭을 사용한 방법이 전처리를 사용하지 않은 방법 보다는 5.47% 향상된 성능을 보여 주었고, 히스토그램 평활화(Histogram equalization)

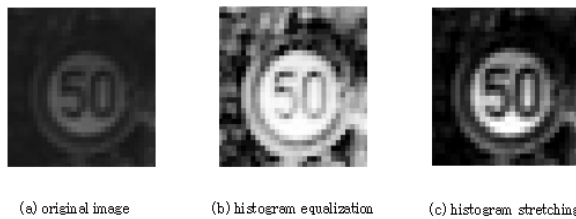


그림 9. 이미지 전처리 결과 비교 (a) 원본 이미지 (b) 히스토그램 평활화 (c) 히스토그램 스트레칭

Fig. 9. Comparison of image pre-processing result (a) original image (b) histogram equalization (c) histogram stretching

를 사용한 방법보다는 11.45% 성능이 향상 된 것을 알 수 있다. 히스토그램 평활화 방법을 사용 했을 때 전체적으로 이미지는 밝아졌다. 하지만 속도 제한 표지판 분류에 중요한 역할을 하는 숫자 부분이 흐릿해져 중요한 정보가 왜곡 또는 손실 되어 전처리를 사용하지 않은 영상보다 오히려 성능이 떨어진 것으로 분석되었다. 그림 9는 이미지 전처리 결과 비교 그림이다.

3. 속도 제한 표지판 성능평가

본 논문에서는 속도 제한 표지판에 해당하는 데이터 13,490장 중 에서 학습 데이터는 12,550장, 테스트 데이터는 940장을 사용 하였으며, 다양한 크기의 영상을 하나의 크기로 정규화 후 CNN의 입력으로 적용하였다. 속도 제한 표지판 인식에 적합한 컨벌루션 필터를 구성하기 위하여 C1 계층에서 컨벌루션 연산 시 사용한 필터는 0~255의 범위를 9개의 영역으로 나누고 이 9개의 영역에서 각각 임의의 값을 추출해 낸 후 그 값의 음수와 양수 값을 이용하여 하나의 필터 값을 구성하였다. 이와 같은 방법으로 300번의 필터 생성 과정을 거쳐 만들어진 필터 중에서 최적의 필터가 선택되면 필터뱅크를 업데이트 하였다. CNN의 구조에서 마지막 S2 계층의 출력은 랜덤 포레스트 분류기의 특징 벡터로 입력된다.

제안하는 2계층 CNN (2CNN)의 성능을 비교하기 위해서, 특징 추출 단계는 제안하는 방법을 그대로 사용하고 분류 단계의 인식기를 다음과 같은 3가지 방법을 적용하여 그 성능을 비교 하였다.

- ① 제안한 알고리즘(2CNN+RF)
- ② Multi-layer Perceptron (2CNN+MLP)
- ③ Support Vector Machine (2CNN+SVM)

성능 평가는 표 1의 테스트 데이터를 모두 사용하였다. 그림 10은 비교 실험을 통한 각 클래스별 Precision율을 보여주고 있다. 그림에서 보는 바와 같이 전체적으로 제안하는 방식은 91.2%의 인식 성능으로 2CNN+SVM에 비해서는 5.2%, 2CNN+MLP에 대해서는 1.3% 높은 성능을 보여 주고 있음을 알 수 있다. 특히, 속도 제한 표지판 중 20, 30,

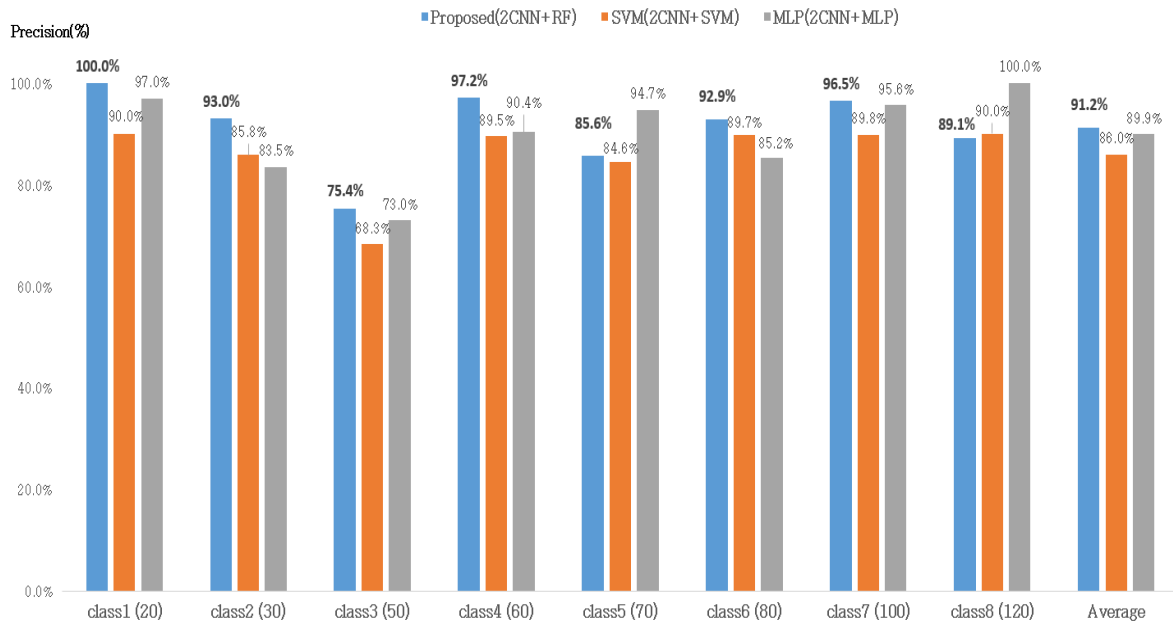


그림 10. 성능 비교 실험 결과
 Fig. 10. Result of performance comparison

50, 60, 80, 100 은 다른 알고리즘 보다 인식 성능이 상대적으로 높은 것을 알 수 있다. 20 속도 제한 표지판의 경우 학습 데이터양이 극히 적었음에도 불구하고 제안된 알고리즘이 100%로 높은 Precision율을 보였다. 그러나, 50 속도 제한 표지판은 테스트 영상에 복잡한 배경이 많이 포함되어 있고 특히 이미지 사이즈를 정규화 하는 과정에서 특징의 손실이 많았기 때문에 75.4%로 낮은 precision율을 보여 주었다. 70과 120 속도 제한 표지판의 경우 모양 특징이 비슷한 20 속도로의 오 분류가 많아 다른 분류기를 사용했을 때보다 낮은 precision율을 보였다. 2CNN+MLP의 경우 전체적으로 제안하는 방법보다 성능이 낮았지만 70과 120 속도 제한 표지판에서는 94.7%와 100%로 제안하는 방법에 비해 높은 성능을 보여 주었다. Penalty(C value)=5, variance(σ^2)=1 인 RBF(Radial Basis Function)를 커널 함수로 사용하는 2CNN+SVM의 경우 총 7개의 클래스에 대해서 낮은 성능을 보여 주었다.

추가적으로 전체 인식 시간에 소요된 시간을 측정하였다. 동일한 시스템 환경에서 1개의 CPU만을 사용했을 때 측정된 결과는 제안한 알고리즘이 1개 영상 당 0.24ms의

시간이 소요 되었고, 2CNN+MLP는 0.3ms, 2CNN+SVM의 경우 2.46ms가 소요되었다. 이 결과, SVM의 연산 시간은 특징벡터의 차원에 매우 민감하며, 실시간 처리에는 적합하지 않음을 알 수 있다. 반면 제안하는 방법은 전체 특징 벡터를 모두 사용하지 않고 랜덤하게 샘플링 하는 방법을 사용함으로써 연산시간이 적게 걸리는 랜덤 포레스트의 특성을 잘 활용하였기 때문에 2CNN+SVM에 비해 매우 우수한 결과를 보여 주었다.

IV. 결론 및 향후 연구 방향

본 논문에서, 낮은 해상도, 속도 제한 표지판의 왜곡, 어두운 조명, 낮은 대비와 같은 환경에서 속도 제한 표지판 인식 성능을 향상하기 위하여, 2계층 Randomly-CNN 구조를 제안하였다. 견고한 속도 제한 표지판 인식을 위하여, 먼저 낮은 해상도와 대비로 인해 흐릿한 영상에 히스토그램 스트레칭을 적용하여 속도 제한 표지판의 숫자와 배경의 구분을 두드러지게 하였다. 또한 전통적인 CNN과 달리

속도 제한 표지판 내 숫자의 좋은 특성을 추출하기 위하여 CNN의 첫 번째 컨벌루션 계층에서 랜덤으로 생성된 필터 중 가장 성능이 좋은 필터를 선택하여 사용하였으며 두 번째 계층에서는 특징 맵 생성 시 서로 다른 가중치를 사용하여 의미 있는 특징이 생성되도록 하였다. 마지막으로 기존의 CNN에서 신경망을 사용하는 것과 달리 적은 학습 데이터로도 일반화 성능이 뛰어난 이진 결정 트리의 앙상블인 랜덤 포레스트를 사용함으로써 높은 특징 차원을 가짐에도 불구하고 빠르고 견고한 인식을 가능하게 하였다.

본 논문의 실험 결과에서 보는 것과 같이 제안한 알고리즘은 낮은 해상도와 어두운 조명을 가지는 환경뿐만 아니라 속도 제한 표지판의 왜곡, 낮은 대비 환경에서도 견고하고 정확한 인식 성능을 보여준다. 다만 몇몇 속도 제한 표지판의 경우 다른 알고리즘에 비해 낮은 것을 볼 수 있지만, 전체적인 성능을 비교했을 때 비교 알고리즘에 비해 높은 인식률을 보여주었다.

본 논문은 기존의 전통적인 필터를 사용하는 대신 랜덤으로 생성된 필터 중 성능이 가장 좋은 필터를 선택함으로써 의미 있는 특징을 추출할 수 있게 되었으며, 또한 신경망 대신 이진 결정 트리의 앙상블인 랜덤 포레스트를 사용함으로써 빠른 수행 속도와 우수한 성능을 보여주었다.

향후 연구에서는 다양한 사이즈 입력이 가능하도록 Spatial Pyramid Pooling(SPP) 적용에 대해 고려해 볼 것이며, 각 계층마다 좀 더 의미 있는 특징이 추출될 수 있도록 컨벌루션 필터의 최적의 가중치를 발견하는 알고리즘을 연구하여 새로운 CNN 구조를 개발할 것이다.

참 고 문 헌 (References)

- [1] L. Kwangyoung, K. Seunggyu and B. Hyeran, "Real-time Traffic Sign Detection using Color and Shape Feature," Korea Computer Congress 2012, Vol 39, No 1. pp. 504-506, June, 2012.
- [2] G. J. L. Lawrence, B. J. Hardy, J. A. Carroll, W. M. S. Donaldson, C. Viskis and D. A. Peel, "A study on the feasibility of measures relating to the protection of pedestrians and other vulnerable road users," Final Tech. Report, TRL. Limited, pp. 206, June, 2004.
- [3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE Con. Computer Vision and Pattern Recognition, vol. 1, pp. 886 - 893, 2005.
- [4] N. Barnes, A. Zelinsky and L. Fletcher, "Real-time speed sign detection using the radial symmetry detector," IEEE Trans. Intelligent Transportation Systems, Vol. 9, No. 2, pp. 322-332, 2008.
- [5] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno and F. Lopez -Ferrerias, "Road-Sign Detection and Recognition Based on Support Vector Machines," IEEE Tran. Intelligent Transportation Systems, Vol. 8, No. 2, pp. 264-278, June, 2007.
- [6] Y. Aoyagi and T. Asakura, "A Study on Traffic Sign Recognition in Scene Image Using Genetic Algorithms and Neural Networks," IEEE Int. Conf. Industrial Electronics, Control, and Instrumentation, Vol. 3, pp. 1838-1843, Aug. 1996.
- [7] G. JaWon, H. MinCheol, K. Byoung Chul and N. Jae-Yeal, "Real-time Speed-Limit Sign Detection and Recognition using Spatial Pyramid Feature and Boosted Random Forest," 12th International Conference on Image Analysis and Recognition, pp.437-445, July, 2015.
- [8] M. Mathias, R. Timofte, R. Benenson and L.V. Gool, "Traffic sign recognition -How far are we from the solution?," IEEE Int. Con. Neural Networks, pp. 1-8, 2013.
- [9] N. Barnes, A. Zelinsky and L.S. Fletcher, "Real-time speed sign detection using the radial symmetry detector," IEEE Trans. Intelligent Transportation Systems, pp. 322-332, 2008.
- [10] Y. Aoyagi and T. Asakura, "Detection and recognition of traffic sign in scene image using genetic algorithms and neural networks.," SICE-ANNUAL CONF. pp. 1343-1348, 1996.
- [11] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in Proceedings of the ACM International Conference on Multimedia, pp. 675-678, November, 2014.
- [12] Y. LeCun, C. Cortes and C.J. Burges, "The MNIST database of handwritten digits," 1998.
- [13] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, pp. 2278-2324, 1998.
- [14] P. Sermanet, Y. LeCun, "Traffic sign recognition with multi-scale convolutional networks," IEEE Int. Conf. Neural Networks , pp. 2809-2813, 2011.
- [15] A. de la Escalera, J. Armingol, J. Pastor and F. Rodriguez, "Visual Sign Information Extraction and Identification by Deformable Models for Intelligent Vehicles," IEEE Tran. Intelligent Transportation Systems, Vol. 5, No. 2, pp. 57-68, June, 2004.
- [16] D.S. Kang, N.C. Griswold and N. Kehtarnavaz, "An invariant traffic sign recognition system based on sequential color processing and geometrical transformation," Proc of IEEE Southwest Symposium on Image Analysis and Interpretation, pp. 88-93, 1994.
- [17] X. Baro, S. Escalera, J. Vitria, O. Pujol and P. Radeva, "Traffic Sign Recognition Using Evolutionary Adaboost Detection and Forest-ECOC Classification," IEEE Trans. Intelligent Transportation Systems, Vol. 10, No. 1, pp. 113-126, Mar, 2009.
- [18] X. Glorot, A. Bordes and Y. Bengio, "Deep sparse rectifier networks", in Proceedings of the 14th International Conference on Artificial

Intelligence and Statistics, Vol. 15, pp. 315-323, 2011.

- [19] K. Kyungmin, H. Jungwoo and Z. Byoungtak, "Character-based Subtitle Generation by Learning of Multimodal Concept Hierarchy from Cartoon Videos," Journal of Korea Intelligent Information System Society, Vol. 42, No. 4, pp. 451-458, 2015.

[20] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5-32, 2001.

[21] J. Stallkamp, M. Schlipsing, J. Salmen and C. Igel, "The German traffic sign recognition benchmark: a multi-class classification competition," IEEE Conf. Neural Networks, pp. 1453-1460, July, 2011.

저 자 소 개



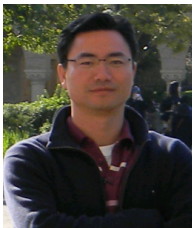
이 은 주

- 2001년 : 계명대학교 컴퓨터공학과 졸업(공학사)
- 2004년 : 계명대학교 컴퓨터공학과 졸업(공학석사)
- 2006년 : 계명대학교 컴퓨터공학과 박사수료
- ORCID : <http://orcid.org/0000-0002-3225-3788>
- 주관심분야 : 컴퓨터 비전 및 패턴인식



남 재 열

- 1983년 : 경북대학교 전자공학과 졸업(공학사)
- 1985년 : 경북대학교 대학원 전자공학(공학석사)
- 1991년 : University of Texas at Arlington 전기공학(공학박사)
- 1985년 5월 ~ 1987년 7월 : 한국전자통신연구소 연구원
- 1991년 9월 ~ 1995년 2월 : 한국전자통신연구소 선임연구원
- 1995년 3월 ~ 현재 : 계명대학교 컴퓨터공학과 교수
- ORCID : <http://orcid.org/0000-0001-7288-283X>
- 주관심분야 : 영상압축, 영상통신, 멀티미디어 시스템



고 병 철

- 1998년 : 경기대학교 전자계산학과 졸업(이학사)
- 2000년 : 연세대학교 대학원 컴퓨터학과 졸업(공학석사)
- 2004년 : 연세대학교 대학원 컴퓨터학과 졸업(공학박사)
- 2004년 : 3월-2005년 8월 삼성전자 통신연구소 책임연구원
- 2005년 : 9월 ~ 현재 : 계명대학교 컴퓨터공학과 교수
- ORCID : <http://orcid.org/0000-0002-7284-0768>
- 주관심분야 : 비전기반 화재감지, 영상검색, 의료영상처리