

# 합성곱 신경망을 이용한 실내 공간에서의 사람 인식

이승수 김만배

강원대학교 컴퓨터정보통신공학과

\*Rugh@kangwon.ac.kr manbae@kangwon.ac.kr

## Human Recognition using CNN under Indoor Environment

Seungsoo Lee and Manbae Kim

Computer & Communications Engineering, Kangwon National University

### 요약

실내공간에서 사람 검출 및 인식은 지능영상정보 시스템에서 중요한 기능중의 하나이다. 실내공간에서 사람검출을 위한 트래킹 과정에서 발생하는 표류(drifting)는 성능 저하의 큰 원인을 제공한다. 표류로 인한 비사람의 트래킹을 방지하기 위해서 사람여부를 결정하는 것이 필요하다. 이를 위해서 합성곱 신경망을 적용한 결과 80%의 사람 인식율을 얻었다.

### 1. 서론

제한된 실내공간에서 카메라로 사람을 검출하는 것은 꾸준히 연구되어 왔다 [1,2]. 사람 검출을 위해서 필요한 기능이 트래킹인데, 트래킹은 객체와 주변 환경과의 색의 유사성, 조명에 의한 조도 변화 등으로 표류(drifting)가 발생하고, 이 표류는 트래킹의 문제점으로 인식되고 있다. 따라서 표류가 발생하면 대부분 일정시간 후에는 정지상태로 간다. 또한 사람도 정지상태로 가기도 한다. 이 정지상태의 트래커가 사람인지 아닌지의 판단 성능이 중요해진다. 즉 정상적으로 특정 객체가 정지상태에 있을 때에 이 객체가 사람인지의 여부를 결정하는 것이 필요한 기능이다.

본 논문에서는 이 기능을 구현하기 위해서 최근 영상 인식 분야에서 사용되고 있는 합성곱 신경망(convolutional neural network)을 활용한다. 이 과정을 통해서 표류 또는 정상 트래킹되는 객체가 사람인지를 결정하게 되어, 전체 점유센서의 성능을 향상시키고자 한다. 트래킹 성능은 본 연구의 관심은 아니고, 주 관심은 트래킹되는 객체의 사람 판별이다. 또한 점유센서는 움직임이 없는 객체를 목표로 해서 사람 판별이 되기 때문에, 트래킹이 수행될 때 정지트래커만 타겟으로 한다.

### 2. 합성곱 신경망의 구조

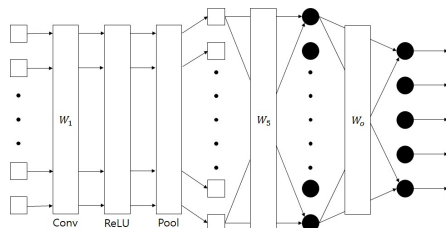


그림 1. 5-class-classification 구성도

제안하는 신경망은 5-class-classification과 2-class-classification 신경망이다. 5-class-classification 신경망은 사물을 구분하는 네 종류의 객

체와 사람으로 총 5가지 class를 분류한다. 2-class classification 신경망은 사람과 비사람으로 총 2가지 class를 분류한다.

그림 1은 5-class-classification 신경망의 구성도를 보여준다. 그림의 신경망은 Convolution layer를 통해 특징맵을 추출한 후 Pooling layer를 통과하여 크기를 축소시킨다. 다음으로 분류 신경망을 통과하여 결과를 출력한다.

입력은 64x64의 흑백 이미지이므로 입력 노드의 개수는 4096이다. Convolution layer는 1개로 구성하고 필터의 크기와 개수를 변경하며 실험한다. Convolution layer의 출력은 ReLU 활성화함수를 통과해 출력된다. 다음으로 1개의 Pooling layer를 배치한다. Pooling 방식은 2x2 mean pooling을 사용한다. 이어지는 분류 신경망은 fully connected network로써 1개의 은닉층과 출력층으로 구성된다. 은닉층에서는 ReLU 활성화함수를 사용하고 은닉 노드의 개수는 50개이다. 분류할 class가 5개이므로 출력 노드는 5개로 구성하고 활성화함수는 Softmax 함수를 사용한다. 표 3은 해당 신경망의 구성을 보여준다.

표 1. 신경망의 구성. fs는 filter size, nof는 number of filter를 의미한다.

layer	feature	activation function
input layer	64 x 64 node	-
convolution layer	fs x fs x nof	ReLU
pooling layer	2 x 2 mean pooling	-
hidden layer	50	ReLU
output layer	5	Softmax

신경망의 가중치는 Xavier 초기화를 사용해 초기화하고 가중치 갱신 방식은 stochastic gradient descent (SGD)를 사용한다. Momentum이나 Dropout 등 다른 기법은 사용하지 않는다.

### 3. 객체 분류의 종류

5-class-classification 신경망이 분류할 class는 총 5가지로 다음과 같다. class 1은 바닥과 벽 등 매끈한 형태를 포함한다. class 2는 특정한

형태인 의자를 포함한다. class 3은 책상과 화이트보드로 각진 형태의 다리가 있는 물건을 포함한다. class 4는 책장, 컴퓨터, 상자 등 육면체의 물건을 포함한다. 마지막으로 class 5는 사람을 포함한다. 2-class classification은 표 2에서 보여진 class type에서 class 1부터 4까지를 class 1로 하고 사람을 class 2로 하여 총 2가지 class를 분류한다. 표 2와 3은 각 신경망의 class type을 보여준다.

표 2. 5-class-classification 신경망의 class type.

class type	object
class 1	바닥, 벽
class 2	의자
class 3	책상, 화이트보드
class 4	책장, 컴퓨터, 상자 등
class 5	사람

표 3. 2-class-classification 신경망의 class type.

class type	object
class 1	비사람
class 2	사람

## 4. 실험 결과

### 1. 5-class-classification

제안방법의 실험은 연구실에 실행하였다. 실험에서는 연구실에서 카메라를 천장에 수직으로 설치하였다. 카메라를 2.7m 높이의 천장에 설치하고 사람은 입출입하게 된다. 영상 해상도는 720x480이고 RGB 입력이다. 사람이 없는 상태에서 촬영한 영상에서 64x64 영상을 sliding window로 획득하고 저장한다. 사람이 입실하면 제안한 트래킹을 실행하면서 트래커를 추적하게 된다. 이 트래커가 정지하면 2장에서 제안한 CNN으로 사람을 판별하게 된다.



그림 2. 5-class-classification 신경망에서 class 1~5의 예시

실험에서 입력 영상은 class별로 각 300장씩 총 1500장을 사용하였다. 신경망의 학습율인  $\alpha$ 는 0.01로 고정하고 convolution filter의 크기와 개수를 조절하며 실험하였다. 입력 영상의 80%는 학습 데이터, 20%는 테스트 데이터로 사용하였고 표 3은 이에 따른 결과를 보여준다.

표 4. 5-class-classification 신경망의 분류 정확도.

epoch	filter		
	3x3x5	5x5x5	7x7x5
10	41.3%	37.6%	41.3%
50	46.6%	43.0%	46.6%
100	51.6%	47.3%	52.3%
200	55.3%	51.0%	47.6%
500	49.6%	50.0%	50.6%
1000	54.6%	49.0%	50.0%

### 2. 2-class-classification



그림 3. 2-class-classification 신경망에서 class 1과 2의 예시

2-class-classification의 입력 영상은 1절의 5-class-classification에서 사용한 것과 동일하다. 하지만 class 1~4를 통합하게 되면 사람을 표현하는 class 2와 개수 차이가 커지므로, class 2의 일부 데이터에 대해 Rotation을 수행하였다.

실험에서 사용한 입력 영상은 class별로 400장씩 총 800장을 사용하였다. 학습율은 1절과 마찬가지로 0.01로 고정하고 convolution filter의 크기와 개수를 조절하며 실험하였다. 표 5는 이에 따른 결과를 보여준다.

표 5. 2-class-classification 신경망의 분류 정확도.

epoch	filter		
	3x3x5	3x3x10	9x9x20
50	77.5%	75.0%	73.7%
100	75.6%	75.6%	76.2%
200	76.8%	76.2%	74.3%
500	76.2%	76.8%	-
1000	78.7%	75.6%	-
2000	75.6%	75.6%	80.0%

## 5. 결론

본 논문에서는 합성곱 신경망을 이용하여 사람 인식을 하는 방법을 제안하였다. 사람인식 성공률이 최대 80%인데, 트래킹에서 발생하는 표류를 제거하는데 큰 도움이 될 것으로 여겨진다.

### 감사의 글

2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2017R1D1A3B03028806)

### 참고 문헌 (References)

[1] Y. Benezeth et al. "Towards a sensor for detecting human presence and characterizing activity", Energy and Buildings, 43, 2011.  
 [2] J. Han and B. Bhanu, "Fusion of color and infrared video for moving human detection", Pattern Recognition, 40, 2007.