

일반논문 (Regular Paper)

방송공학회논문지 제23권 제2호, 2018년 3월 (JBE Vol. 23, No. 2, March 2018)

<https://doi.org/10.5909/JBE.2018.23.2.274>

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

비전 점유센서를 위한 합성곱 신경망 기반 사람 인식

이 승 수^{a)}, 최 창 열^{a)}, 김 만 배^{a)‡}

CNN-based People Recognition for Vision Occupancy Sensors

Seung Soo Lee^{a)}, Changyeol Choi^{a)}, and Manbae Kim^{a)‡}

요 약

대부분의 건물 등에 설치된 점유센서는 PIR(pyroelectric infra-red)이 주로 활용되고 있다. 하지만 PIR은 온도 변화를 감지하는 기능 때문에 정지된 사람을 감지할 수 없는 단점이 있다. 최근 이 단점을 극복하기 위해 카메라 비전 센서의 연구가 진행되고 있다. 비전 센서는 객체 트래킹을 통해 정지된 사람을 검출한다. 그러나 객체 트래킹은 트래커 표류가 발생하는 문제점이 있다. 본 논문에서는 정지 트래커가 사람을 포함하는지의 여부를 판단하기 위하여 합성곱 신경망 기반 사람 인식 기법을 제안한다. 실험에서는 카메라로 획득한 영상에 제안 방법을 적용한 결과 약 88%의 정확도로 사람과 비사람이 분류가 되어 실제 점유센서에 활용이 가능하다는 것을 증명하였다.

Abstract

Most occupancy sensors installed in buildings, households and so forth are pyroelectric infra-red (PIR) sensors. One of disadvantages is that PIR sensor can not detect the stationary person due to its functionality of detecting the variation of thermal temperature. In order to overcome this problem, the utilization of camera vision sensors has gained interests, where object tracking is used for detecting the stationary persons. However, the object tracking has an inherent problem such as tracking drift. Therefore, the recognition of humans in static trackers is an important task. In this paper, we propose a CNN-based human recognition to determine whether a static tracker contains humans. Experimental results validated that human and non-humans are classified with accuracy of about 88% and that the proposed method can be incorporated into practical vision occupancy sensors.

Keywords : occupancy sensor, camera, CNN, people recognition, tracking

a) 강원대학교 컴퓨터정보통신공학과(Department of Computer and Communications Eng., Kangwon National University)

‡ Corresponding Author : 김만배(Manbae Kim)

E-mail: manbae@kangwon.ac.kr

Tel: +82-33-250-6395

ORCID: <http://orcid.org/0000-0002-4702-8276>

※ 이 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2017R1D1A3B03028806).

※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2017R1D1A3B03028806).

· Manuscript received October 19, 2017; Revised January 24, 2018; Accepted March 6, 2018.

Copyright © 2016 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

1. 서론

현재 많은 건물에는 사람이 존재하면 점등하고(light on), 사람이 없으면 소등하는(light off) 점유센서(occupancy sensor, motion sensor)가 설치되어 있다. 전기소비를 줄이기 위해 설치된 것으로 대부분의 점유센서는 현재 PIR(pyroelectric infra-red)을 사용하고 있다^{1,2)}. PIR은 사람의 신체에서 발생하는 열(thermal temperature)을 감지하고, 열의 변화량을 계산하여 모션을 측정한다. 따라서 사람이 정지 상태에 있을 때에는 검출 성능이 낮아진다. 이 센서는 1) 정지된 사람의 검출이 불가능하고, 2) 뜨거운 공기의 흐름 또는 햇볕 복사(radiation)에 민감하고, 3) 사람이 아닌 생물체와 사람을 구분하지 못하는 등의 단점이 있다.

최근에 PIR 대신에 카메라 비전센서를 활용하려는 연구가 진행되고 있다³⁻⁶⁾. 카메라 비전 센서는 위에서 언급한 PIR의 단점을 극복할 수 있고, 부가적으로 사람 추적, 사람 명수 파악, 사람 행위 등 다양한 지능 정보의 획득이 가능하다.

Benezeth 등은 CAPTHOM 프로젝트의 일환으로 비전센서를 활용하여 사람 점유 및 행위를 분석하는 연구를 수행하였다³⁾. 이를 위해 측면에 카메라를 고정하고 얼굴 검출 및 트래킹 기술을 활용하였다. Han 등은 비전 센서와 PIR를 이용하여 사람 검출 연구를 수행하였다⁴⁾. Nakashima 등은 카메라 센서를 활용하여 움직이는 사람을 검출하는 방법을 제안하였다⁵⁾. Amin 등은 PIR과 비전센서를 이용하여 사람의 명수를 구하는 연구를 진행하였다⁶⁾.

상기 비전 점유센서는 PIR의 단점인 정지 객체에 대한 검출 문제를 극복하기 위해서 일반적으로 객체 트래킹을 수행한다. 트래킹은 사람이 정지 상태에 있더라도 사람의 점유를 알 수 있기 때문에 필요한 기능 중의 하나이다.

그러나 트래킹은 객체와 주변 환경과의 색의 유사성, 조명에 의한 조도 변화 등으로 표류(drifting)가 발생하고, 이 표류는 트래킹의 문제점으로 인식되고 있다. 또한 표류가 발생하면 대부분 일정시간 후에는 정지 상태를 유지한다. 또한 사람도 정지 상태로 가기 때문에, 이 트래커가 사람인지 아닌지의 판단 성능이 중요해진다. 그림 1은 트래킹에서 발생하는 정지 트래커의 경계박스들을 보여준다. 청색 경계박스들은 정지된 트래커인데, 그림 1(a)에 있는 한 개의 박스가 트래킹 오류로 발생하는 정지 박스이다. 이 박스는

사람이 존재하지 않으므로, 제거되어야 한다. 또한 그림 1(b)를 보면 4개의 정지 트래커가 존재하는 데, 우측 하단에 있는 트래커는 사람을 포함하고 있지만, 나머지 3개의 트래커는 표류 트래커이므로 제거되어야 한다.

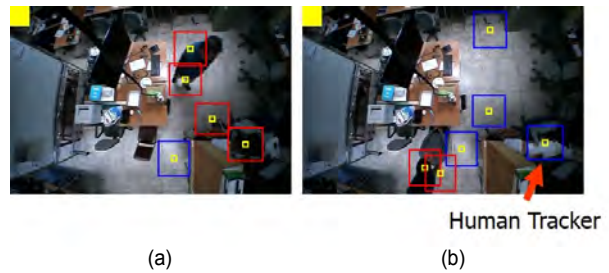


그림 1. 검출된 정지 경계박스들. 정지 트래커는 청색이고, 모션 트래커는 적색임
Fig. 1. The detected stationary bounding boxes marked in blue. The moving boxes are marked in red.

본 논문에서는 이 기능을 구현하기 위해서 최근 영상인식 분야에서 사용되고 있는 합성곱 신경망(convolutional neural network, CNN)을 활용한다. 이 신경망을 통해서 표류된 또는 정상 트래킹된 객체가 사람인지를 결정하게 되어, 점유센서의 성능을 향상시키고자 한다.

입력영상에 합성곱 가중치를 적용하여 분류에 필요한 특징을 생성하고, 영역을 이동하면서 이 과정을 반복한다. 즉 국부 영역으로부터 의미 있는 특징을 생성하는 방식에 해당한다. 검출된 모든 트래커에 대해 합성곱 신경망을 적용하는 것은 큰 처리시간을 필요로 한다. 이동하는 것은 사람일 확률이 높다. 따라서 정지되어 있는 트래커들만 신경망으로 분별하는 것이 타당한 접근 방법이다. 또한 트래킹 성능은 본 연구의 주 관심이 아니고, 핵심은 트래킹되는 객체의 사람 판별이다.

본 논문의 구성은 다음과 같다. 다음 장에서는 제안방법의 기본 개념을 설명하고, III장에서는 객체 트래킹을 설명한다. IV장에서는 CNN을 이용하는 사람 검출 방법을 제안하고, 실험 결과는 V장에서 소개한다. 마지막으로 VI장에서는 결론 및 향후 연구를 정리한다.

II. 제안 방법

제안 방법은 그림 2에서 보여진다. 객체 트래킹은 이동하

는 사람이 정지 상태에 있을 때에도 계속 점유로 판단하기 위해서 필요하다. 트래킹 방법으로 실시간 처리를 위해 운동 히스토리 영상(Motion History Image, MHI) 기반으로 트래킹을 수행한다. 트래커는 이동과 정지를 반복하는데, 이동은 사람일 확률이 크기 때문에, 정지 상태에 있을 때에만 CNN을 이용하여 해당 영상에 사람이 존재하는지를 판단하는 과정에 CNN을 이용한다. 만일 트래커에 사람이 없으면 제거한다.

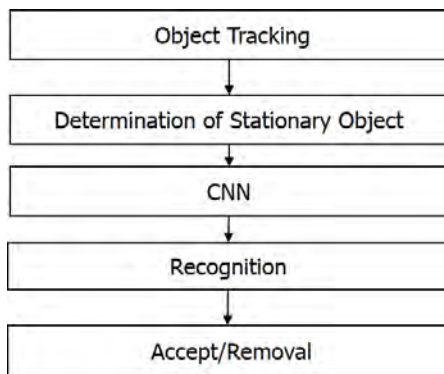


그림 2. 제안 방법의 전체 흐름도
Fig. 2. The overall block diagram of the proposed method

III. 객체 트래킹

본 장에서는 카메라 점유센서의 한 기능인 트래킹을 설명한다. 본 연구의 목적은 트래커에 사람 점유 여부의 판단이기 때문에, 트래킹 기법은 운동 히스토리 영상(motion history image, MHI)을 활용한다.

먼저 영상에서 공간으로 입실하는 사람마다 트래커의 탐색창(search window)을 설정한다. 실내에 설치된 카메라는 천장에 부착한 오버헤드 카메라이다. 천장높이가 보통 3m 정도이므로, 영상에서 사람의 이동은 프레임 간 상당한 거리로 이동하게 된다. 이러한 물리적 현상은 트래커 표류의 원인을 제공한다.

Bovick 등이 제안한 운동정보 표현식인 MHI^[7]는 주로 사람의 행위 및 포즈를 인식하는 행위 검출 분야에서 활발하게 이용되어 왔다. MHI의 기본 정의는 다음과 같다. 식 (1)과 같이 연속영상 I_t, I_{t-1} 로부터 차영상 Δ_t 를 구한다.

$$\Delta_t^i = |I_t^i - I_{t-1}^i| \tag{1}$$

여기서 i 는 픽셀이다. Δ_t 로부터 MHI는 식 (2)와 같이 구한다.

$$MHI_t^i = \begin{cases} \tau & \text{if } \Delta_t^i > T \\ \text{MAX}(0, MHI_{t-1}^i - \delta) & \text{otherwise} \end{cases} \tag{2}$$

윗 식에서 τ 는 사람의 행위가 지속되는 프레임의 개수이고, δ 는 감소(decay) 파라메타이다. $\Delta_t^i > T$ 이면 MHI 값은 τ 가 되고, 아니면 δ 만큼 감소하게 된다. 최근 움직임이 있는 픽셀은 큰 값을 가지는 특성을 가지고 있다. δ 가 크면 픽셀의 MHI는 급격히 감소하게 된다.

실험에서는 $\tau=255$ 를 사용하는데, 픽셀의 변화가 발생하면 MHI=255이다. 반대이면 δ 만큼 감소한다. 그림 3은 연속영상과 MHI의 예를 보여준다. 이 MHI는 트래킹에서 효율적으로 활용된다.

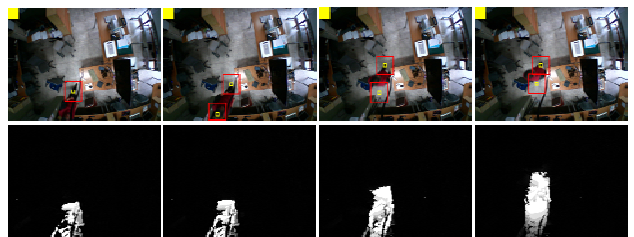


그림 3. 연속영상과 MHI
Fig. 3. Subsequent images and their MHIs

트래커의 경계박스에 대해 픽셀 단위로 전역 탐색을 하면 처리속도가 매우 높다. 속도를 줄이는 방법으로 탐색창을 8x8개의 블록(block)으로 분할하고 (그림 4(b)), 각 블록에는 식 (5)에서 구한 MHI의 에너지 합 E_k 를 저장한다(그림 4(c)). k 는 블록의 번호이다. 그림 4(c)는 그림 4(d)의 MHI 에너지의 관심 영역을 확대한 영상이다.

$$E_k = \sum_{i \in \text{block}(k)} MHI_i \tag{3}$$

이 탐색창 내에서 3x3 블록의 MHI 합이 가장 큰 블록을

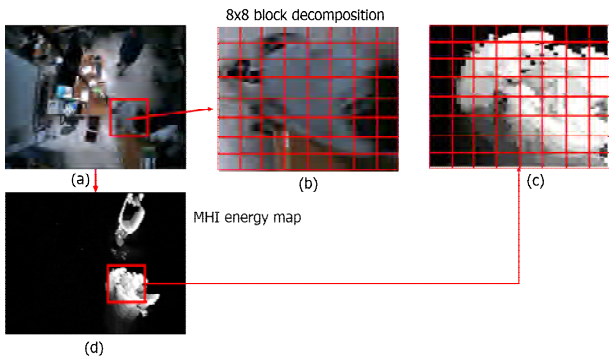


그림 4. 탐색창을 8x8 block으로 분할 후, 블록에 픽셀의 MHI의 합을 저장함
 Fig. 4. Saving the sum of all pixel MHIs into a 8x8 block after decomposition of a search window

찾고, 중심 블록이 트랙커 n 의 추적점 TP_n 이 된다. 정적 객체에 트랙킹 알고리즘을 적용하면 컬러의 변화로 추적점이 변화하게 된다. 사람 판별은 정지 트랙커에만 적용되기 때문에, 트랙커가 정지 상태인지를 조사하고, 정지이면 현재의 추적점을 유지한다. 연속영상 I_t 와 I_{t-1} 을 이용하여 트랙커 n , BB_n 의 차영상을 구하다. 다음에 픽셀의 절대 차이값이 T 보다 작은 픽셀의 개수를 구한 후에, BB 의 픽셀 개수로 나눈 ρ 를 계산한다. ρ 가 0.04 이하이면 정적 트랙커로 판단한다. 이 정적 트랙커는 다음 장의 합성곱 신경망을 이용하여 사람 여부를 판별하게 된다.

그림 5는 제안방법으로 트랙킹한 결과 영상을 보여준다. 각 영상의 MHI는 아래 부분에서 보여진다. 한 사람이 이동



그림 5. MHI 에너지 기반 트랙킹의 결과 영상
 Fig. 5. Resulting tracking images obtained by the MHI energy-based tracking method

하면 두 개의 트랙커가 발생하고, 계속 사람을 추적하는 예를 보여준다. 적색 경계박스는 모션 트랙커이고, 청색은 정지 트랙커이다.

IV. 합성곱 신경망의 구조

합성곱 신경망(CNN)은 일반적인 신경망과 달리 각 노드의 형태가 2차원 필터와 같은 형태로 이루어져 있으며, 퍼셉트론에서는 노드들의 가중치가 단순히 곱해지는 형태이며 각각의 가중치가 입력치와 합성곱이 된다. CNN은 합성곱 계층과 풀링 계층, 완전연결 계층으로 구성된다. 합성곱 계층은 각각의 필터들이 입력영상에 합성곱이 되면서 여러 가지 특징 값들을 추출하고 풀링(pooling)계층에서는 생성된 특징맵을 서브샘플링한다. 최대 풀링, 최소 풀링, 평균 풀링 등이 있다. 실제 분류는 완전연결 계층에서 이루어지며 영상인식의 결과를 출력한다. 가장 마지막 출력 계층은 분류하고자 하는 레이블의 수와 같은 출력을 가지며, 예측되는 레이블 값이 가장 큰 값을 가지게 된다. 분류 단계는 학습을 통해 다양한 경우에 적용할 수 있도록 해주는 것이 목표이며, 기존 신경망과 동일한 구조를 갖는다.

표 1. 5-label classification CNN (L5-CNN)의 레이블 및 해당 객체
 Table 1. Labels and objects for 5-label classification CNN (L5-CNN)

Label	Object
L1	floor, wall
L2	chair
L3	desk, whiteboard
L4	bookshelf, computer, box
L5	human

CNN의 목적은 객체 분류이기 때문에, 먼저 객체 분류에 대해서 설명한다. 레이블 개수는 5 및 2로 설정한다. 먼저 5-label classification 신경망(L5-CNN)이 분류할 레이블은 총 5가지로 표 1에서 보여진다. 레이블 1인 L_1 은 바닥과 벽 등 명암의 변화가 적은 영상을 포함한다. L_2 는 특정한 형태인 의자이다. L_3 은 책상과 화이트보드로 각진 형태의 다리가 있는 물건을 포함한다. 책장, 컴퓨터, 상자 등 육면체의 물건들은 L_4 이다. 마지막으로 L_5 는 사람만을 포함한다(그

림 6). 2-label classification 신경망(L2-CNN)은 이진 클래스이므로 $L_1 \sim L_4$ 는 한 개의 L_1 로 통합하고, L_5 를 L_2 로 한다 (그림 7).

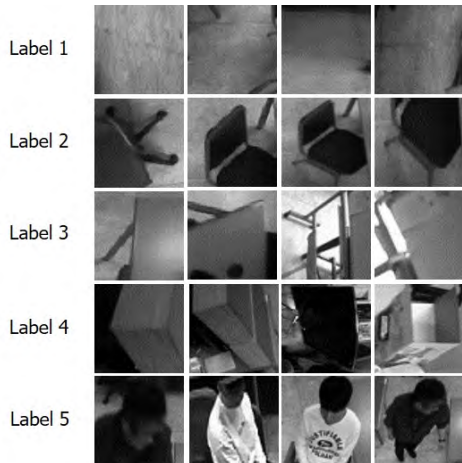


그림 6. L5-CNN의 레이블 1~5 및 해당 학습 영상들
Fig. 6. Training images belonging to [L_1, L_5] in L5-CNN



그림 7. L2-CNN의 이진 레이블 및 해당 학습 영상들
Fig. 7. Training images belonging to [L_1, L_2] in L2-CNN

그림 8은 제안하는 합성곱 신경망의 네트워크 구조를 보

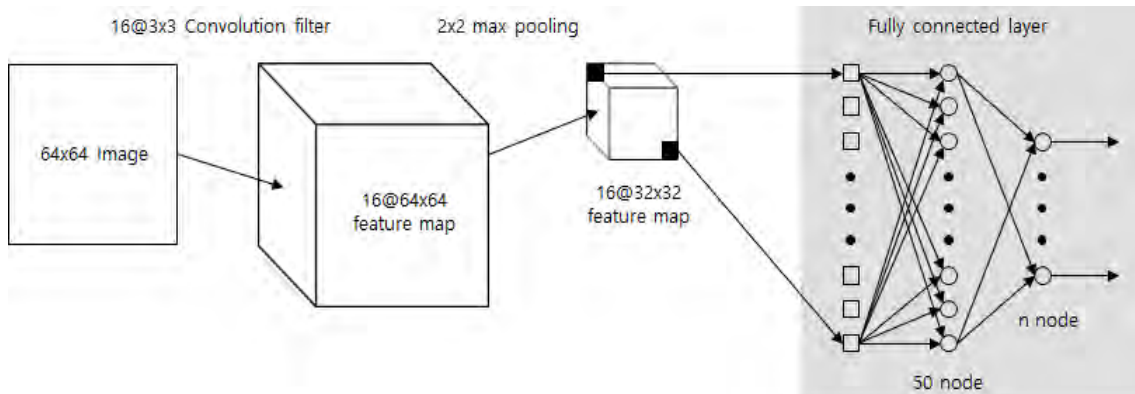


그림 8. 제안하는 신경망의 기본적인 네트워크 구조
Fig. 8. Basic network structure of the proposed neural network

여준다. 입력은 64x64의 grayscale 영상이므로 입력 노드의 개수는 4,096개다. 합성곱 계층은 입력영상의 특징맵(feature map)을 추출하고 ReLU^[8] 활성화함수를 적용하여 풀링 계층으로 전달한다. 풀링 계층은 특징맵의 크기를 축소시킨다. 풀링 방식은 최대 풀링을 사용한다. 이어지는 계층은 분류 신경망으로 은닉 계층에서는 ReLU를 사용한다. 출력 계층에서는 분류할 레이블의 수 n에 따라 노드 개수를 결정하고 활성화함수로는 Softmax 함수를 사용한다. 표 2는 신경망에서 사용하는 파라메타들을 보여준다. 신경망의 가중치 초기화는 Xavier 초기화^[9]를 사용하고, 가중치 갱신 방식은 확률적 경사 하강법(stochastic gradient descent)을 사용한다.

표 2. 신경망에서 사용하는 파라메타 값들
Table 2. Parameter values used in neural networks

Layer		Activation function
input layer	64 x 64	
convolution layer	3 x 3 x 16	ReLU
pooling layer	2 x 2 max pooling	
hidden layer	50	ReLU
output layer	n	Softmax

V. 실험 결과

실험에서 사용한 영상들은 모두 연구실에 획득하였다. 카메라를 2.7m 높이의 천장에 수직으로 설치하고 사람은

입출입하게 된다. 영상 해상도는 720x480이고 RGB 입력이다. 학습과 실험은 Windows10 운영체제에서 MATLAB R2017b를 이용하였다.

이 실험에서 분류의 목적은 영상에 사람이 포함되어 있는지를 판단하는 것이므로, 레이블이 지정된 영상들을 학습에 사용한다. $L_1 \sim L_4$ 의 학습 데이터를 얻기 위해 사람의 비점유 상태에서 촬영한 영상에서 64x64 영상을 이동 윈도우(sliding window)로 획득한다. L_5 의 학습 데이터를 얻기 위하여 다수의 사람을 입실 및 퇴실하게 하면서 사람 학습 영상을 획득하였다. 획득한 학습 영상은 사람이 없는 영상 ($L_1 \sim L_4$)과 사람을 포함한 영상(L_5)이며 약 2,000장이다.

사람이 입실하면 제안한 트랙킹을 실행하면서 트랙커를 추적하게 된다. 이 트랙커가 정지하면 IV장에서 제안한 CNN으로 사람을 판별하게 된다. 연구실에 설치된 카메라로 영상을 입력받아서 III장에서 소개한 트랙킹을 수행하였다. 트랙커가 정지 상태에 있으면 경계박스를 영상으로 저장하였다.

신경망의 학습을 α 는 0.01로 고정하고 합성곱 필터의 수와 은닉 계층의 수를 조절하며 실험하였다. 합성곱 필터의 수는 한 층일 때는 16개이고 이후 한 층씩 늘어남에 따라 2배씩 증가한다. 제로 패딩을 사용하므로 합성곱 연산 과정에서 입력의 크기는 변하지 않는다. 은닉 노드의 수는 한 층일 때는 50개, 두 층일 때는 각각 100개, 50개로 설정하였다.

입력영상의 80%는 학습 데이터, 20%는 테스트 데이터로 사용하였다. 표 3은 L5-CNN의 구조에 따른 분류 정확

표 3. L5-CNN의 구성에 따른 테스트 데이터에 대한 분류 정확도. n-I CNN은 n개의 합성곱 계층, I개의 은닉 계층임

Table 3. Accuracy measured by no. of convolution filters and hidden layers in L5-CNN. n is no. of convolution layers and I is no. of hidden layers in n-I CNN

CNN type	No. of conv layer	No. of hid layer	Accuracy
1-1 CNN	1	1	50.00%
1-2 CNN		2	43.60%
2-1 CNN	2	1	52.80%
2-2 CNN		2	56.80%
3-1 CNN	3	1	60.00%
3-2 CNN		2	50.40%
4-1 CNN	4	1	62.80%
4-2 CNN		2	55.60%

도를 보여준다. 실험한 모든 조건에서 L5-CNN의 분류 정확도는 4-1 CNN에서 최대 62.8%이고, 가장 간단한 구조인 1-1 CNN도 50%의 분류 정확도를 얻고 있다. 이러한 결과는 표류 트랙커를 분류하기엔 L5-CNN은 한계가 있음을 보여준다. 따라서 사람과 비사람을 분류하는 이진 분류기인 L2-CNN을 구현하였다.

L2-CNN의 입력영상은 L5-CNN에서 사용한 것과 동일하다. 하지만 레이블 $L_1 \sim L_4$ 를 L_1 으로 통합하게 되면 사람 레이블인 L_2 와 개수 차이가 커지므로, L_2 의 모든 데이터에 대해 영상 회전 및 크롭(crop)으로 데이터 확장(data augmentation)을 수행하였다. 또한 $L_1 \sim L_4$ 에서 상대적으로 수가 많았던 L_1 과 L_2 의 나머지 영상도 포함하여 새로운 L_1 을 제작하였다. 학습 입력영상은 레이블별로 1,500장씩 총 3,000장을 사용하였다. 표 4는 L2-CNN의 구조에 따른 분류 정확도를 보여준다. 표 3의 L5-CNN보다 분류 정확도가 대폭 향상되는 것을 확인할 수 있다. 가장 간단한 형태인 1-1 CNN의 분류 정확도는 83.17%이고 최대 정확도는 3-2 CNN의 88.17%이다. 네트워크 구조의 관점에서 볼 때, 분류 정확도는 합성곱 계층의 깊이에 비례하여 향상되는데, 반대로 완전연결 계층의 은닉 계층의 수에는 큰 영향을 받지 않는다.

표 4. L2-CNN의 구성에 따른 테스트 데이터에 대한 분류 정확도. n-I CNN은 n개의 합성곱 계층, I개의 은닉 계층임

Table 4. Accuracy measured by no. of convolution filters and hidden layers in L2-CNN. n is no. of convolution layers and I is no. of hidden layers for n-I CNN

CNN type	No. of conv layer	No. of hid layer	Accuracy
1-1 CNN	1	1	83.17%
1-2 CNN		2	84.67%
2-1 CNN	2	1	83.00%
2-2 CNN		2	83.00%
3-1 CNN	3	1	87.17%
3-2 CNN		2	88.17%
4-1 CNN	4	1	86.00%
4-2 CNN		2	86.67%

제안하는 신경망의 성능을 검증하기 위해 SVM(Support Vector Machine)^[12]과 비교 실험을 하였다. SVM은 사람 인

식에서 탁월한 성능을 보여준 기계학습 기법이다. SVM은 이진 분류를 수행하는 L2-SVM과 5가지 레이블을 분류하는 L5-SVM으로 구성하였다. 입력 영상은 제안하는 신경망에서 사용한 것과 동일하게 3,000장을 사용하였다. 성능 비교 결과는 표 5에서 보여진다. 5가지 레이블을 분류하는 경우(표 5) L5-SVM의 분류 정확도는 51.6%이며 이는 1-1 CNN과 1-2 CNN보다 각 1.6%, 8.0% 더 높다. 2-1 CNN부터 4-2 CNN은 L5-SVM보다 최대 9% 높은 성능을 보여준다. 따라서 5 레이블인 경우 적어도 2개의 합성곱 계층을 사용할 필요가 있다.

표 5의 이진 분류에서 L2-SVM은 60.5%의 분류 정확도를 보인다. L2-CNN의 경우 가장 간단한 구조인 1-1 CNN은 83.17%의 분류 정확도를 보이며, 합성곱 계층의 깊이가 증가함에 따라 최대 88.17%까지 성능이 향상된다. 실험 결과는 1-1 CNN과 같이 간단한 구조의 네트워크도 SVM보다 좋은 성능을 보임을 증명한다.

서론에서 언급한 것처럼 활용 측면에서 PIR과 유사한 가격 대비 성능 경쟁력을 갖기 위해서는 점유센서 알고리즘은 단순화되어야 한다. 실용화를 위해서 가장 간단한 구조의 신경망 구조를 설계하고 실제 활용 가능한 성능이 만족되어야 하는데 본 실험 결과는 이것이 가능함을 증명한다.

표 5. CNN과 SVM의 테스트 데이터에 대한 분류 정확도 비교
Table 5. The comparison of classification accuracy for CNN and SVM

5 Labels		2 Labels	
Classifier	Accuracy	Classifier	Accuracy
L5-SVM	51.60%	L2-SVM	60.50%
1-1 CNN	50.00%	1-1 CNN	83.17%
1-2 CNN	43.60%	1-2 CNN	84.67%
2-1 CNN	52.80%	2-1 CNN	83.00%
2-2 CNN	56.80%	2-2 CNN	83.00%
3-1 CNN	60.00%	3-1 CNN	87.17%
3-2 CNN	50.40%	3-2 CNN	88.17%
4-1 CNN	62.80%	4-1 CNN	86.00%
4-2 CNN	55.60%	4-2 CNN	86.67%

마지막으로 분류에 실패한 영상 데이터는 다음 그림 9에서 보여진다. L5-CNN의 오분류된 3장의 영상을 보면, 책장을 바닥으로, 컴퓨터를 의자로, 사람을 의자로 판단한다. L2-CNN에서는 앞의 3장의 영상은 사람인데, 모두 타 객체로 판단한다. 이러한 오분류는 향후 연구에서 개선해야 할 사항이다. 실제 점유센서는 일부 정지 트랙커는 배경영상과의 차이 등을 활용하여 제거할 수 있는 기능이 있다^{[10],[11]}. 이러한 기법과 CNN을 결합하면 보다 우수한 성능의 점유센서를 얻을 수 있다.

Classification Error images							
Label	Ground truth	L ₄	L ₄	L ₅	L ₂	L ₂	L ₂
	Predicted	L ₁	L ₂	L ₂	L ₃	L ₄	L ₅

(a)

Classification Error images							
Label	Ground truth	L ₂	L ₂	L ₂	L ₁	L ₁	L ₁
	Predicted	L ₁	L ₁	L ₁	L ₂	L ₂	L ₂

(b)

그림 9. 분류 오류가 발생한 영상 및 레이블 값. (a) L5-CNN, (b) L2-CNN
Fig. 9. Images with classification error and their labels. (a) L5-CNN and (b) L2-CNN

VI. 결 론

사람 점유를 측정하는 비전센서 기반 점유센서의 중요한 기능중의 하나는 트래킹이다. 트래킹의 단점은 트래킹 오류로 트랙터 표류가 발생하는 것이다. 이 표류된 트랙커가 사람인지 아닌지를 결정해야 하는데, 기존 **hand-crafted feature**, 점유시간을 이용하는 것은 오류가 발생한다. 본 논문에서는 최근 각광을 받고 있는 합성곱 신경망을 이용하여 사람의 여부를 결정하는 방법을 제안하였다. 실험 결과 분류 정확도는 85% 이상이며 가장 간단한 구조의 CNN도 SVM보다도 월등한 성능을 보인다. 이는 제안하는 CNN이 PIR과 유사한 가격 대비 성능 경쟁력을 갖기에 충분함을 증명한다. 제안 방법은 트래킹에서 표류 문제를 해결할 수 있기 때문에 다양한 분야에서 활용이 가능하다.

참 고 문 헌 (References)

- [1] P. Liu et al. "Occupancy inference using pyroelectric infrared sensors through hidden markov model", IEEE Sensors Journal, 16(4), Feb. 2016.
- [2] F. Wahl, M. Milenkovic, and O. Amft, "A distributed PIR-based approach for estimating people count in office environments", IEEE conf on Computational Science and Engineering, 2012.
- [3] Y. Benezeth et al. "Towards a sensor for detecting human presence and characterizing activity", Energy and Buildings, 43, 2011.
- [4] J. Han and B. Bhanu, "Fusion of color and infrared video for moving human detection", Pattern Recognition, 40, 2007.
- [5] S. Nakashima, Y. Kitazono, L. Zhang, and S. Serikawa. "Development of privacy-preserving sensor for person detection," Procedia, 2, 2010.
- [6] I. Amin, A. Taylor, F. Junejom, A. Al-Habaibeh, and R. Parkin, "Automated people-counting by using low-resolution infrared and visual cameras", Measurement, 41, 2008.
- [7] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," IEEE Trans. Pattern Recognition and Pattern Analysis, Vol 23, No. 3, Mar. 2001.
- [8] A. L. Maas, A. Y. Hannun, A. Y. Ng, "Rectifier Nonlinearities Improve Neural Network Acoustic Models," Proc. of the 30th Int. Conf. on Machine Learning, Atlanta, USA, June 2013.
- [9] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep forward neural networks", Int' Conf. Artificial Intelligence and Statistics, Society for Artificial Intelligence and Statistics, 2010.
- [10] J. Gil and M. Kim, "Real-time People Occupancy Detection by Camera Vision Sensor", Journal of Broadcast Engineering, Vol 22, No. 6, Nov. 2017.
- [11] T. Tian, R. Schmidt, H. Liu, A. Hampapur, and M. T. Sun, "Robust Detection of Abandoned and Removed objects in complex surveillance videos," IEEE Trans. Sym. Man and Cybernetics - Part C: Applications and Reviews, Vol. 41, No. 5. pp. 565-577, Sep. 2011.
- [12] C-C. Chang and C-J. Lin, "LIBSVM: A library for support vector machines", ACM Tran. on Intelligent Systems and Technology, Vol 2, No. 3, pp. 27:1-27:27, 2011

저 자 소 개

이 승 수



- 2017년 8월 : 강원대학교 IT대학 컴퓨터정보통신공학과 학사
- 2017년 9월 ~ 현재 : 강원대학교 컴퓨터정보통신공학과 석사과정
- 주관심분야 : 점유센서, 딥러닝, 머신러닝

최 창 열



- 1979년 : 경북대학교 전자공학과 학사
- 1981년 : 경북대학교 전자공학과 석사
- 1995년 : 서울대학교 컴퓨터공학과 공학박사
- 1984년 ~ 1996년 : ETRI 컴퓨터연구단 책임연구원 / 연구실장
- 2009년 ~ 2011년 : 강원대학교 IT대학 학장
- 1996년 ~ 현재 : 강원대학교 컴퓨터정보통신공학과 교수
- ORCID : <http://orcid.org/0000-0002-8340-4195>
- 주관심분야 : 모바일전송, 3D데이터처리, 미디어서비스

저 자 소 개



김 만 배

- 1983년 : 한양대학교 전자공학과 학사
- 1986년 : University of Washington, Seattle 전기공학과 공학석사
- 1992년 : University of Washington, Seattle 전기공학과 공학박사
- 1992년 ~ 1998년 : 삼성종합기술원 수석연구원
- 1998년 ~ 현재 : 강원대학교 IT대학 컴퓨터정보통신공학과 교수
- 2016년 ~ 현재 : 강원대학교 정보통신연구소 소장
- ORCID : <http://orcid.org/0000-0002-4702-8276>
- 주관심분야 : 3D영상처리, 비전점유센서, 객체트래킹