

# 서베일런스에서 회선 신경망 기술을 이용한 사람 추적 기법

강성관\*, 천상훈\*\*

인하대학교 정보공학과\*, 인천재능대학교 정보통신과\*\*

## Human Tracking Technology using Convolutional Neural Network in Visual Surveillance

Sung-Kwan Kang\*, Sang-Hun Chun\*\*

HCI Lab., Department of Computer and Information Engineering, Inha University\*

Department of Information and Technology, Incheon JEI University\*\*

요약 본 논문에서는 현재와 이전의 영상 프레임 뿐 만 아니라 영상의 축척과 이전 위치에 주어진 객체의 비율과 위치 추정에 대한 학습 문제로서 사람 추적 문제를 다룬다. 본 논문에서는 회선 신경망 분류기를 이용한 사람 검출 방법을 제안한다. 제안하는 방법은 신경망을 정규화하고 검출 작업을 위한 특징 표현을 자동으로 최적화함으로써 사람 검출의 정확성을 향상시킨다. 제안하는 방법에서는 감시 영상 시스템에서 실시간 영상이 들어오면 제일 먼저 위치를 추정하는 작업을 수행하기 위하여 회선신경망을 학습시킨다. 기존의 다른 학습 방법과 달리 회선신경망은 두 쌍의 연속된 영상 프레임으로부터 공간적이고 시간적인 특징을 모두 공동으로 학습시킨다. 회선 신경망에 의해 학습된 특징을 이용하는 SVM 분류기의 정확성은 회선 신경망의 정확성과 일치한다. 이것은 자동적으로 최적화된 특징의 중요성을 확인시켜 준다. 그러나, 회선 신경망을 이용한 사람 객체의 분류에 대한 계산 시간은 사용된 특징의 타입과 관계없이 SVM의 것보다 약 40분의 1 정도로 작다.

주제어 : 보행자 추적, 사람 추적, 회선 신경망, 객체 추적, 딥 러닝

**Abstract** In this paper, we have studied tracking as a training stage of considering the position and the scale of a person given its previous position, scale, as well as next and forward image fraction. Unlike other learning methods, CNN is thereby learning combines both time and spatial features from the image for the two consecutive frames. We introduce multiple path ways in CNN to better fuse local and global information. A creative shift-variant CNN architecture is designed so as to alleviate the drift problem when the distracting objects are similar to the target in cluttered environment. Furthermore, we employ CNNs to estimate the scale through the accurate localization of some key points. These techniques are object-independent so that the proposed method can be applied to track other types of object. The capability of the tracker of handling complex situations is demonstrated in many testing sequences. The accuracy of the SVM classifier using the features learnt by the CNN is equivalent to the accuracy of the CNN. This fact confirms the importance of automatically optimized features. However, the computation time for the classification of a person using the convolutional neural network classifier is less than approximately 1/40 of the SVM computation time, regardless of the type of the used features.

**Key Words** : Pedestrian Tracking, Human Tracking, Convolution Neural Network, Object Tracking, Deep Learning

Received 30 December 2016, Revised 23 January 2017  
Accepted 20 February 2017, Published 28 February 2017  
Corresponding Author: Sung-Kwan Kang(HCI Lab., Department of Computer and Information Engineering, Inha University)  
Email: kskk1111@empas.com

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. 서론

실시간 감시 영상에서 침입탐지는 특정 지역의 접근 보안을 위하여 사람의 영상정보를 이용한 사람 검출 및 추적 시스템을 개발하는 것으로 발전하고 있다. 그러한 시스템에서 신뢰성 있는 사람 검출은 영상정보를 사용하는 모든 요소의 기반이 되며, 해당 정보를 이용한 여러 가지 응용에 사용되어지므로, 정확한 사람 객체의 검출 및 추적이 필요하다. 이러한 이유로, 현재 사람 영상 정보를 이용한 보안에 많은 필요성이 증가되고, 국내외적으로 수요가 급증하고 있는 추세이다. 따라서, 실시간 영상에서 이동성에 강인한 새로운 사람 객체의 모델링 및 검출 및 추적 기술이 절실히 요구되고 있다. 또한 고도의 영상처리 기법인 다해상도 사람 검출 및 추적 기술을 이용한 특정 공간내의 물리적 접근 제어 및 사용자 이용 확인 등의 기술 확보가 필요한 시점 있다.

현재의 어플리케이션들은 단순한 움직임 검출을 이용하여 접근 제어 및 보안 분야에 응용이 주로 되고 있다. 이러한 시스템은 중요한 감시 대상에 관계없는 움직임, 조명의 변화, 불필요한 미세 변화 등에 시스템 구동으로 인한 사용, 관리, 저장 공간 등의 문제가 있다.

기존의 연구들에서 다양한 방법들을 통하여 실시간 영상에서 이동하는 사람을 검출 및 추적하는 기법을 개발하였으나, 검출 환경이나 검출 대상의 이동 가변성에 대한 신뢰도가 떨어지는 문제점이 있었다[1][2]. 또한, 현재의 대부분 침입 탐지 시스템은 조명의 변화 및 외부 요인에 의한 잡음 영상에 취약점을 가지며, 이에 따른 물리적 접근 시장 분야에서 매우 난항을 겪고 있다. 현재 DVR 과 CCTV 등의 광범위한 보급으로 인하여 특정 공간내의 출입자 관리를 제공하는 시스템의 완성도가 높히 요구되고 있으며 이를 위하여 새로운 사람 객체의 모델링 및 검출, 추적 등의 기법이 절실히 필요한 실정이다.

객체 추적은 컴퓨터 비전에서 근본적인 문제이다. 색상 기반 방법과 동작 주체 기반과 같은 기존의 특징 기반 방법들은 추적 대상의 단순한 모델을 유지시키고 모델을 과도하게 적응시킴에 의하여 객체 추적을 수행한다 [3][4]. 그러나, 일상생활에서 벌어지는 실제 상황은 이러한 방법들이 추적에 있어서 많은 오류를 발생시키게 하는 단점을 갖는다. 그러한 단점 발생에 대한 첫 번째 이유는 과도하게 적응시키는 객체 모델은 객체의 원래 영

상으로부터 유도될 수 있고, 두 번째 이유는 특징 기반 방법들이 관심 영역 안에 있는 객체를 다른 객체들과 구별하도록 하는 차별적인 모델을 갖지 못하기 때문이다.

본 논문에서 제안하는 방법의 주요한 장점은 현재 실시간 감시 영상 시스템에 있어서 가장 중요시 되는 객체의 배경 영상의 변화 및 객체의 이동에 영향을 받는 추적 대상 객체에 대한 영상 검출 및 추적 성능의 감소 문제를 회선 신경망 기술을 이용하여 개선하고 강인한 객체 추적을 지속시키는 것이다[5].

본 논문의 구성은 다음과 같다. 2장에서 감시 영상 시스템에서 사람 검출과 관련된 기존의 연구에 대하여 서술한다. 3장에서 본 논문에서 제안하는 추적 알고리즘에 대하여 서술하고 추적 시스템의 단일 프레임 정확도를 설명한다. 4장에서 제안하는 회선 신경망을 이용한 사람 추적 시스템과 다른 방법들과의 비교 실험 결과를 보여 주고 5장에서 본 논문에서 제안하는 방법 및 시스템에 대한 향후 연구 방향에 대하여 설명하며, 결론을 맺는다.

## 2. 문제 정립

### 2.1 실시간 영상에서 사람의 특징점 추출 및 검출

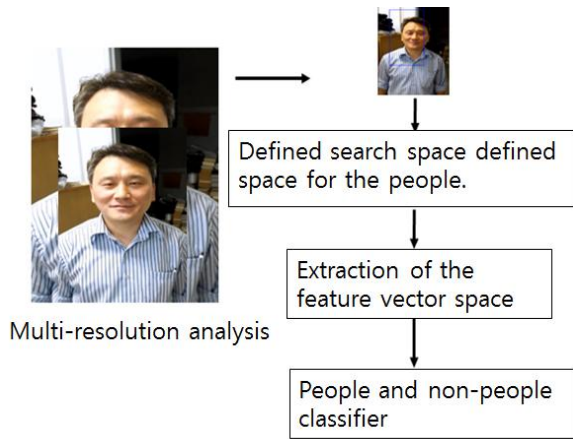
사람 검출은 시공간을 따라서 추적 대상을 검출하여, 침입 탐지 시스템과 시설물/문화재 감시, 범죄 예방 등 실시간 경고 메시지를 필요로 하는 여러 분야에서 사용되고 있다. 개인의 안전을 보장받고 더 많은 물리적 자산과 정보 자산을 보호하기 위하여 CCTV 등을 이용한 보안 감시 시스템의 응용 분야가 늘어나고 있다. 이러한 사회적 추세에 발맞추어 다양한 CCTV 등을 이용한 사람을 검출하고 추적하기 위한 다양한 연구들이 진행되어 오고 있다.

따라서 본장에서는 기존에 연구되었던 사람 검출 방법을 설명하고 이동-가변적인 사람 객체, 방해물이 있는 배경 등에 강인한 결과를 보이는 다양한 사람 검출 연구들을 분석하여 문제점 및 개선 방향을 제시한다. 그리고 본 논문에서 제안하는 방법 및 사람 추적 시스템의 구조에 대하여 설명한다.

실시간 영상에서 사람을 검출하고 안정적으로 추적하기 위해서는 사람과 다른 객체들과의 구별이 명확해야 한다는 환경 조건이 필요하다. 이는 사람 추적 기법에서

만드시 필요한 요소로 여겨지는 것이다. 기존 연구들에서는 실시간 입력 영상으로부터 주위 환경 변화에 영향을 덜 받도록 전처리 과정을 수행하며, 그 결과로 추출된 사람의 특징적인 데이터 요소로부터 수치화된 특징 벡터 값을 얻어낸다. 이러한 최적화된 전처리 과정은 사람의 특징데이터를 뽑아내는 속도를 상당히 증가시킨다. 이러한 시스템에서는 특징점 검출 단계를 크게 학습 단계와 검출 단계로 구성한다[6].

외형기반 검출 방법은 사람의 픽셀 단위의 값을 이용하는 것이 아니라, 사람 영역인지 아닌지를 판별하기 위한 영역 기반의 사람 검출 방법을 의미한다. 외형기반 사람 검출 시스템의 일반적인 방법은 [Fig. 1]과 같다.



[Fig. 1] Appearance-based Detection Model Person Pyramid

외형 기반 검출 방식에서 기초가 되는 사람 검색 창 크기를 정의할 필요가 있으며, 정의된 검색창은 실제 영상에서 존재하는 사람의 크기와 다를 수 있다. 따라서 [Fig. 1]에서와 같이 사람을 찾고자 하는 원이미지의 사이즈를 변경시켜가며, 검색하기 위한 검색창을 쉼운다. 검색창내의 영역에 사람이 포함될 수도 있고, 포함되지 않을 수도 있으며, 이 영역이 사람인지, 비사람인지를 판별하는 알고리즘을 수행함으로써 사람을 검출한다. 이러한 기법을 다중해상도 기법이라고 한다.

또한 신경망 시뮬레이터로 거의 폭넓게 사용된 회선 신경망의 구현의 부족함 등이 있었는데 이러한 것들은 다른 분류기에 비해서 회선 신경망을 조직적으로 연구하는 그룹들이 적었기 때문이다[7]. 따라서, 본 논문에서는 최근 사람 검출 시스템에서 가장 범용적인 분류기인

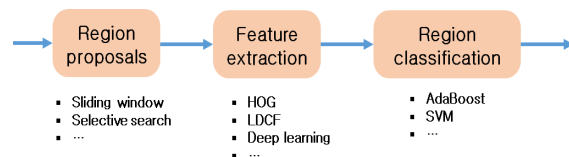
SVM과 잘 제어된 회선 신경망의 성능 비교 결과를 제안한다.

본 논문에서는 실시간 영상에 대하여 회선 신경망 기술이 기존의 SVM과 비교하여 더 우수한 검출 정확성을 나타낸다는 것을 보인다. 회선 신경망 기술이 기존의 연구들에서 정면 사람 검출에 대하여 성공적으로 수행되었는데, 이것은 변화가 심한 자연 환경에서 윤곽선 정보를 필요로 하는 객체 검출을 위한 응용에 대한 뛰어난 기술이다. 본 논문에서는 학습 과정에 있어서 회선 신경망과 SVM의 두드러진 차이점을 적용함에 의하여 회선 신경망의 일반화 능력을 향상시키는 방법을 제안하고 회선 신경망에 의해 추출된 특징을 가지고 SVM을 훈련시킴으로써 회선 신경망과 SVM의 본질적 결합을 평가한다. 이러한 융합적 SVM의 정확성이 기본 회선 신경망의 정확성보다 더 높을 지라도 그것은 단지 정규화 된 회선 신경망의 정확성과 거의 일치한다[8].

### 2.2 사람 검출 파이프 라인

최근 사람 검출을 위한 많은 서로 다른 방법이 제안되었고, 상업적인 응용과 군사용 응용 모두에 성공적으로 구현되었다. 그러한 방법들이 사람 모양을 검출하고 의미론적 표현을 얻기 위하여 원래 데이터를 처리하는 방식이 아주 다름에도 불구하고, 그 방법들은 데이터 처리를 위하여 유사한 파이프라인을 공유한다. 그러한 파이프라인의 입력은 영상의 픽셀 레벨의 표현인 원시 데이터이고, 반면에 출력은 분석된 프레임 안에서 식별되어 온 각각의 사람과 일치하는 서로 다른 크기로 된 영역 상자의 집합으로 이루어진다. 그러한 파이프라인은 [Fig. 2]와 같이 세 가지 주요한 단계로 구성된다[9]. 그 단계는 다음과 같다.

- (단계 1) 영역 도출
- (단계 2) 특징 추출
- (단계 3) 영역 분류



[Fig. 2] Common Pipeline for Human Detection

첫 번째 단계인 영역 도출에서, 전체 프레임은 잠재적으로 사람을 포함하고 있는 영상의 일부인 후보 영역을 추출하기 위하여 분석되어진다. 그러한 단계의 입력은 전체 프레임인 반면에 출력은 이질적인 차원과 비율을 가진 영역의 집합이다. 슬라이딩 윈도우 방법은 가장 단순한 영역 도출 알고리즘의 실례이고 다중 축척과 종횡 비에서 영역을 추출하도록 하기 위하여 적용될 수 있다.

두 번째 단계인 특징 추출 단계에 대해서는 서로 다른 방법들이 다수 제안되어져 왔다. 그러한 방법들은 아주 다르게 데이터를 처리하고 입력 데이터의 다양한 비선형 변환과 수집된 기울기, 지역적 채도 대비와 같은 이질적인 시각적 특징을 이용한다.

마지막으로, 분류 단계는 사람 모양과 일치하는 후보들의 집합 안에 있는 영역들을 식별하는 것을 목표로 한다. 분류기는 주어진 영역과 관련된 특징 벡터를 입력으로 받아들이고, 전형적으로 그러한 영역이 사람을 포함하는 영역인지 아닌지를 가리키는 이진 레이블을 산출한다.

### 3. 사람 추적 알고리즘의 설계

#### 3.1 이동 가변적인 회선신경망을 이용한 사람 검출기

회선 신경망(Convolutional neural network) 기술은 LeCun에 의하여 제안되었는데 이 기술에서는 특징 추출기와 분류기를 동등하게 다룬다[10]. 기존의 학습 기반 방법과 검출 방법에 의한 추적의 주요 문제는 추적의 잘못된 연결을 초래하는 거짓 긍정 결합(false positive match)이다[11]. 그 이유는 그러한 방법들이 추적 대상 주위에 있는 모든 가능한 영역에서 객체 검출기나 외형 모델을 적용하는 것을 기반으로 하고 현재 프레임에서의 객체 검출은 이전 프레임에 있는 추적 대상의 위치 식별에 의존하지 않기 때문이다. 따라서, 많은 사람들이 몰려 있는 실시간 영상에서 추적에 방해가 되는 다른 객체들이 추적 대상과 유사할 때, 객체 검출기는 추적 대상과 아마도 이동 문제를 유발하는 방해 객체 모두에 대하여 비슷한 높은 검출 점수를 생성할 것이다. 예를 들어, 추적기가 군중 속에 있는 한 사람의 머리를 추적하기 원한다면, 군중 속에 있는 다른 사람들의 머리가 또한 좋은 결

합이 될 수 있기 때문에 머리 검출기를 사용한다면 추적은 매우 어렵다. [Fig. 3]에서 추적 대상이 이전 프레임에서  $O$  에 있다고 가정한다. 위치  $A$ 와  $A'$ 이 객체 검출기에 의해 비슷한 높은 검출 점수를 생성할 수 있다는 것은 분명하다. 왜냐하면, 검출기는 추적 대상이 이전 프레임에서  $O$  에 있다는 것을 알지 못하기 때문이다.

이러한 이동 문제를 완화시키기 위하여, 제안하는 방법에서는 이전 프레임에 있는 객체의 위치와 외형이 현재 프레임에 있는 객체를 검출하는 것을 돕는다고 간주한다. 검출에 대한 이러한 추가적인 정보를 이용함으로써, 제안하는 방법에서는 현재와 이전의 영상 프레임 뿐만 아니라 크기와 이전 위치에서 주어진 추적 대상의 비율과 위치 추정의 능력이 있는 추적기로써 객체 검출기를 효과적으로 전환시킨다. 본 논문에서 제안하는 사람의 이동 가변성에 강인한 추적 시스템을 구현하기 위한 기본 학습기로써 회선신경망을 이용한다[10]. 회선신경망은 지역화 된 시각적 특징을 추출할 수 있다고 증명되어져 왔고 다양한 시각적인 인식 어플리케이션에 폭넓게 사용되어져 왔기 때문이다[11]. 완전히 연결된 신경망과는 다르게 회선신경망은 영상이 강한 2차원 지역적 구조를 가진다는 사실을 기반으로 지역화 되는 은닉 단위들의 수용 필드를 제한함으로써 효과적으로 지역적 특징을 추출할 수 있는 장점이 있기 때문이다.



**Previous image**      **Current image**

[Fig. 3] ability problem that occurs when an object interferes be similar to the target track.

추적에 있어서 기존의 회선신경망 적용의 방해 요소는 인식이나 검출 작업에는 적합하지만 추적 작업에는 부적당한 이동-불변적인 구조를 가지고 있다는 것이다 [11]. 본 논문에서는 이동-가변적인 구조를 가지는 회선

신경망 추적기를 설계하고 구현한다. 그러한 구조는 회선신경망 모델이 검출기로부터 추적기로 전환되도록 하는 핵심 역할을 수행한다. 구현하는 시스템에서 사람에 대한 특징은 오프라인 훈련 동안에 학습되어진다.

### 3.2 사람 검출을 위한 회선신경망의 최적화

사람 검출의 상황에서 회선신경망의 이용은 최근에 이루어졌고 객체 인식 분야에서 상당히 뛰어난 결과를 보이고 있다. 본 논문에서는 회선 신경망의 모든 단계와 제안하는 파이프라인을 설명한다. 학습의 경우에 작은 세목이 정확성의 관점에서 좋은 결과를 도출하는 열쇠가 된다. 파이프라인의 각 단계를 주의깊게 분석하고 최적화함으로써 제안하는 방법에서는 수작업화 된 특징을 기반으로 하는 기존의 방법들에 비해 추적 성능을 10% 정도 향상시킨다.

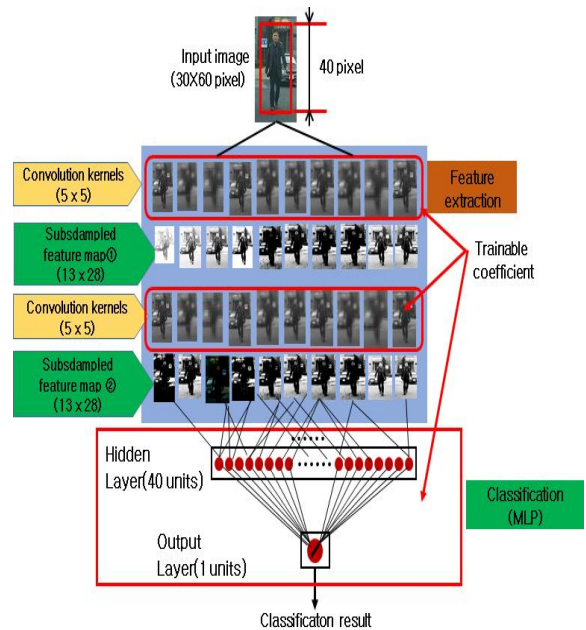
제안하는 방법에서는 검출 파이프라인의 첫 번째 단계는 사람을 나타낼 수 있는 후보 영역을 식별하는데 있다. 이러한 단계는 계산적 효율성과 정확한 인식 작업을 위한 열쇠이다. 반면에, 부정적인 영역의 대부분을 효율적으로 버림으로써 다음 단계에 입력으로 넣어줄 영역의 수는 1/3 수준으로 감소될 수 있다. 이것은 특징 추출의 계산 시간이 많이 소요되는 회선신경망에 요구되어질 때 최고의 중요성이 된다. 반면에, 이 알고리즘은 많은 긍정적인 영역을 버리지 않음으로써 시스템의 전체적인 정확성에 미치는 영향을 최소화시킨다.

### 3.3 회선 신경망

(CNN, Convolutional Neural Network)

회선 신경망은 첫 번째 레이어가 계층적 특징 추출기로서 동작하도록 구성되어진 다층 퍼셉트론의 특별한 변형이다[11]. 보통의 완전 연결형 다층 퍼셉트론과의 차이는 특징 추출 레이어(또는 특징 맵)안에 있는 각 처리 노드가 선행 레이어의 다른 서브영역으로 연결되어진다는 것과 각 특징 맵 안에 있는 처리 노드들이 똑같은 가중치 벡터를 공유한다는 것이다[12]. 이러한 연결 구조는 본질적으로 선행 레이어의 출력결과 상에서 학습 가능한 회선 필터링을 수행하기 위하여 특징 맵을 구성한다. 특징 맵은 보통 차원수를 감소시키고 작은 왜곡에 대한 강인함을 향상시키기 위하여 서브 샘플링 레이어 다음에 나오게 된다. 더 높은 레벨의 특징 맵은 포유류 영상 시스

템 안에 있는 것과 닮은 계층적 특징 모델을 구현하기 위하여 몇 가지 더 낮은 레벨의 맵으로부터 그것들의 입력을 취할 수 있다. 회선 신경망의 마지막 레이어는 이전의 레이어들에 의해 추출된 특징들 상에서 일반적 목적의 분류기를 구현하기 위하여 완전 연결되어진다. 회선 신경망의 구조와 중간 처리 결과는 [Fig. 4]에서 보여주는 것과 같이 실험에서 사용된다.

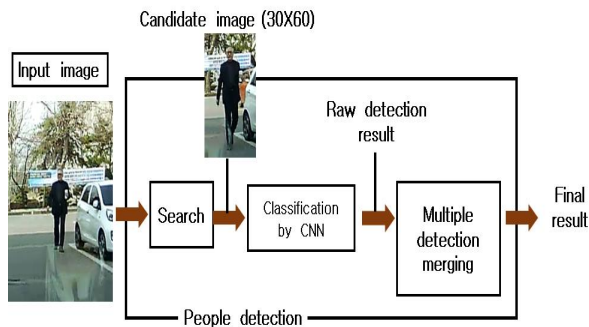


[Fig. 4] Structure and intermediate processing results of convolutional neural network

입력 레이어의 크기는 30 x 60 픽셀이다. 입력 영상 안에 있는 사람의 높이는 학습하는 동안에 40 픽셀이다. 위와 아래에서 10 픽셀의 비교적 큰 여백은 회선 필터링의 두 단계 동안에 이진 효과를 보상하기 위하여 필요하다[13]. 첫 번째와 두 번째 레벨에 10개의 특징 맵이 있다. 회선과 서브샘플링은 [9]에서 설명된 것과 같이 계산적 효율성을 위하여 하나의 단계에서 구현된다. 각각의 두 번째 레벨 특징 맵은 정확히 하나의 첫 번째 레벨 특징 맵과 연결되어진다. 여기서 첫 번째 레벨 특징 맵은 어떠한 계층적 특징 결합도 수행되지 않는다. 모든 레벨 2의 특징 맵과 완전히 연결된 숨겨진 레이어의 크기는 40 단위이다. 제안하는 방법에서는 고정된 학습률을 가진 확률 변화도 기반 학습 알고리즘을 이용하는 교차-엔트로피 에러 함수로써 회선 신경망을 학습시킨다[14].

### 3.4 제안하는 시스템의 설계 및 구현

[Fig. 5]는 본 논문에서 제안하는 사람 검출 시스템의 블록다이어그램을 나타낸다. 카메라로부터 입력된 영상은 C. Papageorgiou의 시스템과 유사하게 몇 가지 해상도에서 30 x 60 픽셀의 검출 윈도우로써 탐색되어진다 [10]. 걸음걸이 크기는 양방향 모두에서 3 픽셀이고 제안하는 시스템에서는 다단계 영상을 만들기 위하여 재조정 인자(rescaling factor) 0.9를 이용한다.



[Fig. 5] The block diagram of the proposed human detection system

30 x 60 후보 영상은 조명 보상과 같은 전처리 과정 없이 회선 신경망의 분류기로 입력되어진다. 회선 신경망은 각 후보 영상에 대하여 분류 점수를 부여하고 임계값보다 더 높은 점수를 가진 후보 영상들은 원시 검출 리스트(raw-detection list)에 저장된다[15]. 원시 검출 리스트는 분류 점수에 의해 정렬되어진다. 회선 신경망 분류기가 작은 이동과 크기 변화에 강인하기 때문에 보통 각각의 사람은 원시 검출 리스트에 몇 가지 결과를 생성할 것이다. 제안하는 방법에서는 동일한 사람에 대한 중복된 결과를 제거하기 위하여 최종 검출 결과를 산출하기 전에 중복 검출에 대하여 합병 작업을 수행한다.

## 4. 실험 결과

이 절에서는 제안하는 검출 시스템의 분류기를 평가하는 실험적 결과를 제안한다. 실험에서는 회선 신경망에 대한 최대 마진 학습의 중요성을 평가하고 또한 서로 다른 특징 표현을 이용하는 SVM 분류기와 비교한다. 본 논문에서는 다양하게 변화하는 실시간 동영상에 대하여 제안하는 회선신경망 추적기에 대하여 실험을 진행하였

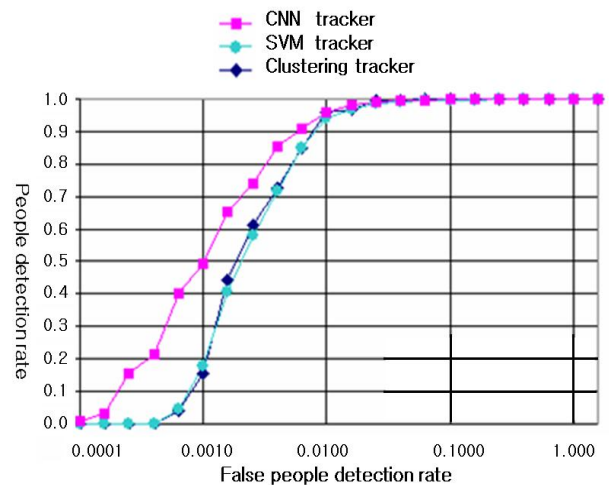
다. 실험은 서로 다른 환경하에서 강인함과 알고리즘의 몇 가지 요소의 성능을 평가하기 위하여 설계된 두 부분으로 나뉘어진다. 실험에서 보여지는 테스트 과정은 학습되지 않았다. 테스트 과정은 연구실 안과 밖에서 모두 진행하였으며 학습 영상은 서로 다르다. 그러므로, 본 논문에서 제안하는 방법에서는 회선 신경망에 의하여 선택된 특징들이 실험을 통하여 폭 넓게 응용될 수 있다는 것을 증명한다. 제안하는 알고리즘은 C++, C#언어를 이용하여 구현되었고 Intel i7 4.0 GHz 컴퓨터에서 테스트하였다. 코드 최적화 없이 프로그램은 평균 640 x 480 해상도에서 초당 10~15 프레임을 실행한다.

회선 신경망 학습은 먼저 특징 추출 네트워크와 분류 네트워크를 학습하고 학습된 특징 추출 네트워크를 이용해 위치 추정 네트워크를 학습하는 순서로 진행한다. 먼저 한 번의 학습에 사용되는 샘플의 개수는 너무 적을 경우 과적합(overfitting) 문제가 발생하므로 64개로 하였다.

똑같은 사람이 학습과 테스트 모두에서 나타나지 않도록 하기 위하여 학습과 테스트 데이터의 시간 범위를 기록하는 것 사이에 겹쳐짐은 없다. 서로 다른 데이터 집합의 크기는 <Table 1>에서 보여준다.

<Table 1> The size of the different data sets

Data Set	Positive Sample	Negative Sample
Training Data	460	856
Test Data	195	856



[Fig. 6] Comparison of the human detection rate for convolutional neural network and other trackers

[Fig. 6]의 알고리즘별 사람 객체 검출에서 제안하는 회선 신경망 기술과 다른 방법을 사용하였을 때의 실험 결과를 보면 제안하는 방법의 거짓 사람 검출율이 SVM 이나 클러스터링 방법을 이용한 추적기보다 더 낮다는 것을 알 수 있다.

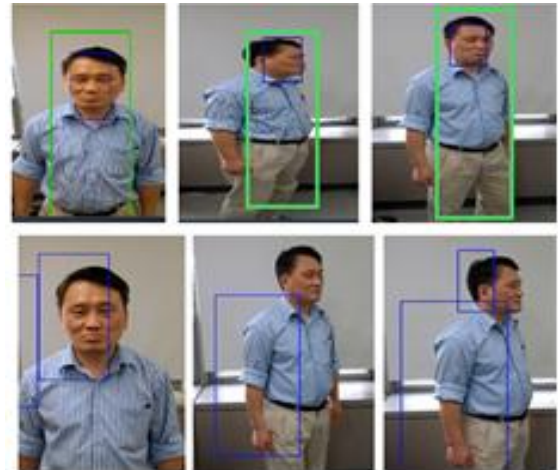
<Table 2> Object detection rate using convolutional neural network classifier

	# of Image	Success	Fault	Success Rate(%)	Average search time (seconds)
Object 1	100	98	2	98	0.012048
Object 2	100	96	4	96	0.145631
Object 3	100	98	2	98	0.009524
Object 4	100	100	0	100	0.019417
Object 5	100	96	4	96	0.009709
Object 6	100	95	5	95	0.019048
Object 7	100	98	2	98	0.009709
Object 8	100	97	3	97	0.015021
Object 9	100	98	2	98	0.016598
Object 10	100	99	1	99	0.009935
Average	100	97.5	2.5	97.5	0.026664

<Table 2>에서는 제안하는 방법인 회선 신경망 분류기를 이용하여 각 객체별로 검출 성공률 및 평균 검색 시간을 나타내고 있다. 10명의 서로 다른 사람에 대하여 검출 실험을 하였으며 10명 모두에 대하여 평균 97.5%의 비슷한 검출율을 보이고 있다.

제안하는 방법에서는 제어된 설정 안에서 일반적인 SVM과 회선 신경망 분류기를 비교한다. 가우시안 커널 SVM은 Torch 라이브러리를 이용하여 학습되어진다. 학습 오류와 마진 사이의 가우시안 커널의 변화와 절충 인자는 각각의 실험에 대해서 별도로 최적화되어진다. 결과로 나오는 support 벡터의 수는 5,000의 범위 안에 있고 서로 다른 특징 파라미터 설정 사이에 중요하게 변화하지 않는다.

본 실험에서는[Fig. 7] 에서와 같이 추적대상자가 큰 비율과 시야 변화에 놓일 때 추적 성능을 보여준다. 회선 신경망 추적기는 학습 기반 방법을 가진 SVM 추적기와 비교된다. 본 실험에서는 오프라인 학습된 지원 벡터 기계가 추적에 대한 분류기로서 적용되는 지원 벡터 추적기와 회선신경망 추적기를 비교하였다.



[Fig. 7] Tracking people in the video there is a change of magnification and field of view

다음 <Table 3>에 있는 실험 결과에서는 회선신경망을 이용한 추적기, SVM을 이용한 추적기, 클러스터링에 의한 추적기에 대하여 평균 추적 성공률과 에러율을 보여준다. 각 추적기에 대한 평균적인 추적 성공률은 회선 신경망을 이용한 추적기가 94.5% 이고 SVM을 이용한 추적기가 90.7%, 클러스터링에 의한 추적기가 89.7%를 나타내었다.

<Table 3> Track success rate and error rate in each tracker

Tracker	Average Tracking Success Rate (%)	Error Rate (%)
Clustering Tracker	89.7	10.2
		10.8
SVM Tracker	90.7	6.3
		6.7
CNN Tracker	94.5	4.6
		4.0

### 5. 결론

본 논문에서는 사람 검출 시스템에서 분류기로서 회선 신경망의 사용을 제안하고 평가하였다. 실험 결과로부터 가장 중요한 결론은 검출 작업에 자동적으로 최적화된 좋은 특징들이 높은 정확도의 검출을 위하여 필수적이라는 것이다. 회선 신경망의 거짓 사람 검출율은 표준 Haar 웨이블릿과 SVM 결합의 거짓 사람 검출율의

1/5보다 작다. 그러나, SVM 분류기의 정확도는 똑같은 특징들을 사용했을 때 기본 회선 신경망의 정확도를 능가하는데 그것은 SVM에 고유한 최대 마진 학습을 사용하기 때문이다. 회선 신경망이 큰 차이의 결정 경계를 얻기 위하여 학습 동안에 규칙화 될 때 거짓 사람 검출율(FP-rate)은 기본 회선 신경망의 거짓 검출율의 절반으로 떨어진다. 이것은 사실상 회선 신경망 특징과 SVM 결합에 해당되는 결과이다. 게다가 회선 신경망의 또 하나의 큰 장점인 최적의 특징을 학습하기 위한 능력은 아주 낮은 계산 비용이다. 회선 신경망으로써 하나의 사람 후보 영역을 처리하기 위한 시간은 SVM에 의해 필요로 하는 시간의 3% 보다 적다.

회선 신경망이 더 높은 검출 정확도와 상당히 낮은 계산 시간을 필요로 하기 때문에 사람 검출에 대한 분류기으로써 범용적으로 이용되는 수동적으로 설계된 특징을 이용하는 SVM 결합 보다 더 정확한 추적을 위한 사람 분류기에 더 적합하다.

## REFERENCES

- [1] Y. Li, H. Ai, T. Yamashita, S. Lao, and M. Kawade, "Tracking in low frame rate video: A cascade particle filter with discriminative observers of different life spans," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 30, No. 10, pp. 1728 - 1740, Oct. 2008.
- [2] D. Ramanan, D. A. Forsyth, and A. Zisserman, "Tracking people by learning their appearance," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 29, No. 1, pp. 65 - 81, Jan. 2007.
- [3] B. Wu and R. Nevatia, "Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors," *Int. J. Comput. Vis.*, Vol. 75, No. 2, pp. 247 - 266, Nov. 2007.
- [4] C. Papageorgiou, T. Evgeniou, and T. Poggio, "A trainable pedestrian detection system," in *Proc. Intelligent Vehicle Symposium IV'98*, Stuttgart, Germany, Oct. 1998.
- [5] D. Valentin, H. Abdi, A. J. Otoole, and G. W. Cottrell, "Connectionist Models of Face Processing: A Survey", *Pattern Recognition*, Vol. 27, pp. 1209-1230, 1994.
- [6] L. Zhao and C. Thorpe, "Stereo and neural network-based pedestrian detection," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 1, No. 3, pp. 148 - 154, Sept. 2000.
- [7] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vis.*, Vol. 77, No. 1 - 3, pp. 125 - 141, May 2008.
- [8] P. Viola, M. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Proc. IEEE International Conference on Computer Vision, ICCV 2003*, Nice, France, Oct. 2003.
- [9] A. Shashua, Y. Gdalyahu, and G. Hayun, "Pedestrian detection for driving assistance systems: Single-frame classification and system level performance," in *Proc. IEEE Intelligent Vehicle Symposium, IV 2004*, Parma, Italy, June 2004.
- [10] A. Broggi, A. Fascioli, M. Carletti, T. Graf, and M. Meinecke, "A ultiresolution approach for infrared vision-based pedestrian detection," in *Proc. IEEE Intelligent Vehicle Symposium, IV 2004*, Parma, Italy, June 2004.
- [11] J. Fan, M. Yang, and Y. Wu, "A bi-subspace model for robust visual tracking," in *Proc. IEEE Int. Conf. Image Process.*, San Diego, CA, pp. 2660 - 2663, Oct. 2008.
- [12] Zhai. Yujia, "Stable Tracking Control to a Non-linear Process Via Neural Network Model", *International Conference on Convergence Technology*, Vol. 5, No. 4, pp.163-169, 2014.
- [13] Nipon. Theera-Umpon, Lee. Sanghyuk, "Similarity Measure Design on High Dimensional Data", *International Conference on Convergence Technology*, Vol. 4, No. 1, pp.43-48, 2013.
- [14] Sunghyuck Hong, "New Authentication Methods based on User's Behavior Big Data Analysis on Cloud", *Journal of Convergence Society for SMB*, Vol. 6, No. 4, pp.31-36, 2016.
- [15] Hyung-Song Shin, Kyun-Tak Kim, Kyu-Jin Lee, Kye-San Lee, "A study on Scalable Video Coding



Signals Transmission using inter-layer Differential OVSF code allocation scheme in MC-CDMA”, Journal of Convergence Society for SMB, Vol. 6, No. 3, pp.49-55, 2016.

강 성 관(Kang, Sung Kwan)



- 2001년 2월 : 인하대학교 컴퓨터공학부 (학사)
- 2005년 8월 : 인하대학교 정보통신공학과 (석사)
- 2006년 9월 ~ 현재 : 인하대학교 컴퓨터정보공학부(박사과정)
- 관심분야 : 컴퓨터 비전, HCI
- E-Mail : kskk1111@empas.com

천 상 훈(Chun, Sang Hun)



- 1990년 2월 : 인하대학교 전자공학과 (학사)
- 1992년 2월 : KAIST 전기및전자공학과 (석사)
- 2000년 2월 : 인하대학교 전자공학과 (박사)
- 2000년 3월 ~ 현재 : 인천재능대학교 정보통신과 부교수
- 관심분야 : 트래픽 제어, 네트워크, 영상 통신
- E-Mail : wintari@naver.com