

Efficient Swimmer Detection Algorithm using CNN-based SVM

Dasol Hong*, Yoon Kim**

Abstract

In this paper, we propose a CNN-based swimmer detection algorithm. Every year, water safety accidents have been occurred frequently, and accordingly, intelligent video surveillance systems are being developed to prevent accidents. Intelligent video surveillance system is a real-time system that detects objects which users want to do. It classifies or detects objects in real-time using algorithms such as GMM (Gaussian Mixture Model), HOG (Histogram of Oriented Gradients), and SVM (Support Vector Machine). However, HOG has a problem that it cannot accurately detect the swimmer in a complex and dynamic environment such as a beach. In other words, there are many false positives that detect swimmers as waves and false negatives that detect waves as swimmers. To solve this problem, in this paper, we propose a swimmer detection algorithm using CNN (Convolutional Neural Network), specialized for small object sizes, in order to detect dynamic objects and swimmers more accurately and efficiently in complex environment. The proposed CNN sets the size of the input image and the size of the filter used in the convolution operation according to the size of objects. In addition, the aspect ratio of the input is adjusted according to the ratio of detected objects. As a result, experimental results show that the proposed CNN-based swimmer detection method performs better than conventional techniques.

▶ Keyword: Object detection, HOG, SVM, CNN

I. Introduction

해수욕장 방문자 수가 해마다 증가함에 따라 물놀이 안전사고 또한 매년 빈번히 발생하고 있다. [Fig. 1]은 국민안전처에서 제공하는 2012년부터 2015년까지의 물놀이 안전사고 횟수를 나타낸다 [1]. 그림을 보면, 물놀이 사고는 지속적으로 발생하며 사고의 대부분이 사망사고인 것을 알 수 있다. 이렇게 해마다 지속적으로 발생하는 물놀이 사고를 방지하기 위해 최근 많은 수의 전문 인력과 장비들이 해수욕장에 투입되고 있다. 뿐만 아니라, 물놀이 안전사고의 사전 예방 체계를 구축하기 위해 지능형 영상 감시 시스템(Intelligent Surveillance System)이 많이 개발되고 있다[2]. 지능

형 영상 감시 시스템은 감시형 CCTV(Closed-Circuit Television)와 IT(Information Technology) 기술을 접목한 방법으로, CCTV에서 얻은 영상을 이용하여 자동으로 사물이나 사람과 같은 객체를 인식하고 추적한다.

최근, 최유정 등[2]은 GMM(Gaussian Mixture Model)[3], HOG(Histogram of Oriented Gradients)[4], SVM(Support Vector Machine)[5]을 사용하는 입수자 검출 알고리즘을 제안하였다. 제안한 알고리즘은 먼저 GMM으로 배경 영상에 대한 모델링과 전경을 추출한다. 그리고 HOG와 SVM으로 검출

• First Author: Dasol Hong, Corresponding Author: Yoon Kim

*Dasol Hong (hda8297@kangwon.ac.kr), Dept. of Computer and Communications Engineering, Kangwon National University

**Yoon Kim (yooni@kangwon.ac.kr), Dept. of Computer and Communications Engineering, Kangwon National University

• Received: 2017. 11. 22, Revised: 2017. 11. 30, Accepted: 2017. 12. 09.

• This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No: R-20160906-004163, Developing Bigdata Autotagging and Tag-based DaaS System)

• This study has been worked with the support of a research grant of Kangwon National University in 2017.

된 객체에 대한 특징 벡터를 추출하여 입수자를 검출한다. 그런데, [2]의 알고리즘에서 사용한 HOG는 독특한 윤곽선을 가진 객체 검출에는 효과적이지만, 객체의 엣지(edge)가 유사하거나 뭉개진 경우 정확한 특징 벡터를 얻을 수 없는 문제가 있다. 따라서, 본 논문에서는 입수자 검출에서 HOG로 발생하는 문제점을 완화하기 위해 이미지의 엣지가 아닌 추상적인 특징 값을 추출하여 사용하는 CNN(Convolutional Neural Network)[6] 기반 입수자 검출 알고리즘을 제안한다. 또한, 본 논문에서는 거리가 먼 곳에서 나타나는 객체들을 효율적으로 검출하기 위해 작은 객체 사이즈에 특화된 CNN 구조를 제안한다.

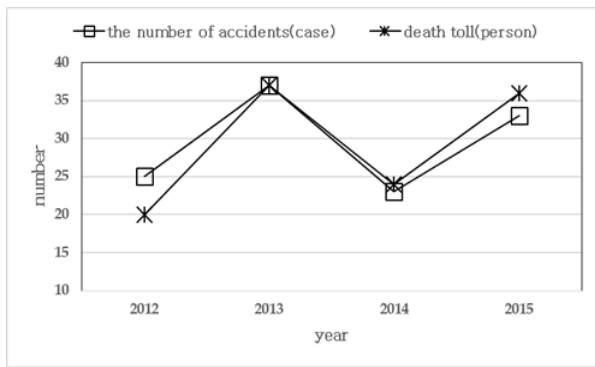


Fig. 1. The status of water safety accident

본 논문에서 제안하는 입수자 검출 알고리즘의 동작은 다음과 같다. 먼저, CCTV로부터 얻은 입력 영상에 GMM을 사용하여 배경에 대한 모델을 만들고, 전경을 추출해낸다. 다음으로 GMM으로 추출된 객체 후보들에 대해 CNN을 사용하여 특징 벡터를 추출한 후 이를 분류에 적용하여 파도와 같은 사물과 입수자를 탐지한다. 검출된 입수자에 대해서는 실시간/지속적인 탐지를 위해 트래커(tracker)를 사용하여 추적한다. 실제 해상용장 데이터를 사용하여 실험한 결과, 본 논문에서 제안한 CNN을 사용한 입수자 검출이 기존 알고리즘보다 성능을 크게 향상시켰다. 또한, CNN의 분류기(classifier)를 softmax 분류기[7]와 kernel svm[5]를 사용하여 각각 실험한 결과 정확도(accuracy) 측면에서 kernel svm가 softmax 분류기 보다 근소한 차이로 좋은 결과를 보였다.

본 논문의 구성은 다음과 같다. 먼저 2절에서는 객체 탐지와 CNN에 대한 기존의 연구들을 분석한다. 다음으로 3절에서는 기존 입수자 검출 알고리즘의 문제를 분석하고 이에 따른 해결책을 서술한다. 그리고 4절에서는 제안한 입수자 검출 알고리즘에 대한 동작 과정과 알고리즘의 각 단계를 자세하게 서술한다. 다음으로 5절에서는 제안한 알고리즘의 성능 평가를 진행하며 기존 알고리즘과의 비교, 분류기의 종류에 따른 성능을 비교, 평가한다. 마지막으로 6절에서는 결론을 맺고 향후 연구에 대해 논의한다.

II. Related works

객체 탐지(object detection)는 이미지나 영상에서 특정 물체를 검출하는 방법으로, 이미지 매칭, 영상 처리, 객체 분류 등 많은 분야에서 응용되어 사용된다. 이러한 응용 분야들 중에서도 특히 감시 카메라를 사용한 영상 처리는 그 용도에 따라 보행자 검출[8], 얼굴 인식[9], 번호판 인식[10], 입수자 검출[2] 등 많은 응용 연구들이 존재한다. 이러한 객체 탐지 분야에서 가장 중요한 알고리즘은 특징 기술자(feature descriptor)[11]이다. 전통적으로, 보행자 검출에서는 HOG, 얼굴 검출에서는 Haar-like feature[12]가 주로 사용되었다. 하지만 이러한 기술자들은 성능을 향상시키는데 한계가 있었으며, 이를 극복하기 위해 최근 새로운 특징 기술자의 연구가 많이 진행되었다. 특히, Felzenszwalb 등[13]이 제안한 Deformable Part Model은 HOG와 LSVM(Linear SVM)[5]을 이용한 기술자로, 보행자 검출에서 많이 사용되었던 알고리즘이다. 하지만, CNN의 상용화와 함께 대부분의 객체 탐지 알고리즘에서 CNN이 사용되고 있으며 최근에는 다양한 구조를 가진 모델들을 활용한 객체 탐지 알고리즘이 등장하고 있다[14-16].

CNN은 영상에 대한 특징 추출과 해당 특징들을 인공 신경망으로 분류해내는 하나의 네트워크 구조이다[14]. CNN은 특징 벡터 추출로 컨볼루션 레이어(convolutional layer)[14]를 사용하는데, 이는 여러 개의 필터(filter)들이 쌓인 구조를 갖는다. 여기서 필터란, 기존 영상 처리 분야에서 사용하는 필터와 유사한 개념이지만 학습을 통해 가중치 값이 조절된다는 차이가 있다. 해당 필터를 이용하는 컨볼루션 레이어의 가장 큰 특징 중 하나는 다양한 필터 연산을 통해 초기엔 지역적인(local) 정보들이 추출되고 최종적으로는 전역적인(global) 정보가 추출된다는 것이다. 이런 컨볼루션 레이어의 특징으로 인해 기존 특징 기술자들을 뛰어넘는 성능을 보이고 있으며, 최근에 다양한 컴퓨터 비전 분야[11-13]에서 폭발적으로 사용되고 있다.

일반적으로, CNN은 컨볼루션 레이어와 softmax 분류기를 사용하지만 softmax 분류기 대신 다른 분류기를 사용하는 방법들이 많이 연구되었다. 먼저, Tang[7]은 softmax 분류기 대신 선형(linear) SVM을 사용해 성능 향상을 시켰으며, Huang 등[17]은 Gaussian kernel SVM을 사용하여 성능을 향상시켰다. 본 논문에서는 이를 응용하여 입수자 검출 알고리즘에서 사용하는 CNN의 분류기에 softmax 분류기와 kernel-SVM을 각각 적용하여 객체를 탐지한다.

III. Research motives

본 절에서는 기존 HOG를 사용하는 입수자 검출 알고리즘의 문제점을 분석한다. 기존 알고리즘에서 HOG를 사용한 대표적인 이유는 HOG가 보행자 검출에서는 사람의 실루엣을 특징 벡터로 매우 효과적으로 표현하여 높은 객체 탐지 성능을 보이기

때문이다. 하지만, 보행자 검출과는 달리, 입수자 검출에서는 HOG가 보행자 검출만큼의 성능을 나타내지 못한다. 즉, HOG가 해수욕장을 배경으로 하는 영상 데이터에서 입수자를 빗 반사, 물결과 같은 동적 요소라고 잘못 인식하는 긍정 오류(false positive)와 동적 요소를 입수자라고 잘못 판단하는 부정 오류(false negative)가 빈번히 발생한다. 이를 자세히 설명하기 위해 먼저 기존 입수자 검출 방법을 설명한다.

해수욕장과 같은 물놀이 영상 데이터에서 입수자를 검출하는 방법은 크게 특징 벡터 추출과 특징 벡터를 통한 객체 분류로 나뉜다. 먼저, 특징 벡터 추출은 입력 영상으로부터 GMM과 HOG를 이용해 사람, 사물과 같은 객체를 벡터로 추출하는 과정이다. 그리고 객체 분류는 추출한 벡터를 기준에 학습시킨 SVM을 통하여 입력 영상에서 파도, 빗 반사, 사람 객체 등으로 분류하여 탐지하는 과정이다. 여기서 SVM은 입력으로 들어오는 특징 벡터가 주로 되어 분류를 진행하기 때문에, 식 (1)과 같은 계산을 통해 특징 벡터간의 유사도를 계산하여 동적 요소와 비슷한 특징 벡터를 갖는 사람 객체의 영상을 얻을 수 있다.

$$y = \sum_{i=0}^n |h[i] - o[i]| \quad (1)$$

식 (1)에서 h 는 사람 객체의 특징 벡터, o 는 사람으로 판단된 동적 요소의 특징 벡터이며 n 은 벡터의 차원 수이다. 즉, 유사도 y 가 낮으면 두 벡터간의 유사성이 높다는 것을 의미하므로, 유사성이 높은 벡터의 영상을 통해 오류가 발생하는 이유를 분석할 수 있다. [Fig. 2]는 식(1)을 사용하여 파도, 빗 반사 객체와 높은 유사성을 나타내는 입수자 객체를 추출해 낸 영상이며 [Fig. 3]은 SVM을 통해 사람 객체로 판단되는 파도와 빗 반사를 보여준다. [Fig. 2]를 보면, 동적 요소와 유사한 벡터 값을 갖는 객체는 대부분 영상 내에 파도를 포함하고 있거나 엷지가 흐릿하다. 그런데, [Fig. 2-(a)]와 같이 파도가 객체와 함께 추출되는 이유는 GMM을 이용한 전경 추출을 진행하기 때문에, 움직임이 겹치게 되면 하나의 객체로 묶여 추출되기 때문이다. 그렇기 때문에 엷지 기반의 특징 벡터 추출 방식을 사용한다면, 영상 내에 입수자가 존재함에도 불구하고 파도로 판단될 확률이 높아진다. 또한, 엷지가 흐릿한 영상은 크기 보정(resize)할 때 사람 객체 이외에 존재하던 잡음 값이 커지게 되고 이 잡음들이 파도와 유사한 엷지 값을 갖게 되는 문제점이 있다.

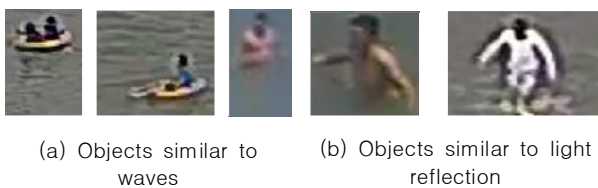


Fig. 2. Objects with a high similarity to the dynamic background

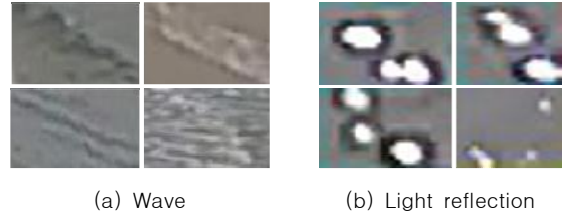


Fig. 3. Elements of the ocean extracted as object candidates

빛 반사의 경우도 빛 반사의 엷지가 사람 또는 튜브와 비슷하기 때문에 HOG를 이용해 뽑아낸 특징 벡터로 빛 반사와 사람을 구별해내는 것은 매우 어렵다. 따라서, 본 논문에서는 위와 같은 문제점들을 해결하기 위해 객체의 엷지가 아닌 추상적인 특징 값을 추출하는 CNN을 사용하여 효율적으로 입수자를 검출하는 알고리즘을 제안한다.

IV. The Proposed Scheme

본 논문에서는 HOG 사용으로 발생하는 긍정 오류와 부정 오류 문제를 완화하기 위해 CNN을 활용한 입수자 검출 알고리즘을 제안한다. 제안하는 CNN 기반 입수자 검출 알고리즘은 HOG 대신 CNN을 사용함으로써, 객체 후보로 파도가 추출된 경우, 객체의 엷지가 흐릿한 경우, 빛 반사와 엷지가 비슷한 경우 등에 대해 효과적으로 객체를 분류 및 탐지한다.

[Fig. 4]는 제안하는 알고리즘의 전체 동작 과정을 나타낸다. 제안하는 알고리즘은 크게 객체 추출, 특징 벡터 추출, 객체 분류, 입수자 추적의 4 단계로 구성된다. 본 절에서는 각 단계에 대한 알고리즘을 상세하게 기술한다.

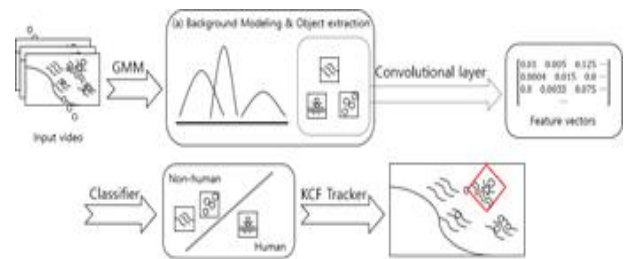


Fig. 4. The operation process of the swimmer detection algorithm using convolutional layer

먼저, 객체 추출 단계에서는 CCTV에서 입력 받은 영상에 GMM을 이용해 배경 모델링을 하고 객체를 추출한다. GMM은 처음으로 각 픽셀마다 가우시안 모델을 여러 개 만든다. 이 때 가우시안 모델은 확률 밀도 함수이며 상황에 따라 그 개수를 임의로 지정할 수 있다. 임의로 지정한 M 개의 확률 밀도 함수들의 선형 결합으로 전체 확률 밀도 함수를 정의할 수 있으며, 다음 식 (2)로 표현된다.

$$p(x|X_T, BG + FG) = \sum_{m=1}^M \pi_m N(x, \overrightarrow{\mu}_m, \sigma_m^2 I) \quad (2)$$

식 (2)는 어느 한 시점 T에서 한 픽셀의 가우시안 혼합 분포를 나타낸 식이다. X는 한 픽셀 값이며 X_T 는 입력 영상, BG는 배경 영상일 확률, FG는 전경 영상일 확률이다. 또한, π_m 은 m 번째 가우시안 분포의 가중치 값이며 $N(x, \overrightarrow{\mu}_m, \sigma_m^2 I)$ 는 평균 $\overrightarrow{\mu}_m$ 과 분산 σ_m^2 을 가지는 정규분포를 의미한다. 여기서 새로운 데이터 $x^{(t)}$ 가 들어온다면, 아래의 식을 통해 가우시안 분포의 가중치 값, 평균, 분산이 재귀적으로 갱신된다.

$$\pi_m \leftarrow \pi_m + \alpha(o_m^{(t)} - \pi_m) \quad (3)$$

$$\overrightarrow{\mu}_m \leftarrow \overrightarrow{\mu}_m + o_m^{(t)}(\alpha/\pi_m)\overrightarrow{\delta}_m \quad (4)$$

$$\sigma_m^2 \leftarrow \sigma_m^2 + o_m^{(t)}(\alpha/\pi_m)(\overrightarrow{\delta}_m^T \overrightarrow{\delta}_m - \sigma_m^2) \quad (5)$$

위 식에서 α 는 오래된 데이터의 영향력을 제한하기 위해 사용되는 상수이며, $\overrightarrow{\delta}_m^T \overrightarrow{\delta}_m$ 는 공분산이다. 그리고 $o_m^{(t)}$ 는 새로운 데이터가 들어오면 가장 큰 값을 갖는 요소에 대해서는 값이 1로 설정되며 나머지는 0으로 설정된다. 최종적으로 GMM은 위의 식들을 이용해 픽셀 별로 가우시안 모델들을 생성하고 이를 이용해 각 픽셀마다 배경 혹은 전경으로 분류해준다. GMM을 이용해 추출한 입수자 후보들은 실제 입수자 이외에도 다양한 객체들이 검출되는데, 이는 GMM으로 바다의 동적인 배경을 완벽하게 표현할 수 없기 때문이다.

다음으로 특징 벡터 추출 단계에서는 컨볼루션 레이어를 이용해 이전 단계에서 추출된 객체 후보자들에 대한 특징 벡터를 추출한다. 그리고 객체 분류 단계에서는 추출된 특징 벡터로 최종적인 입수자를 얻기 위해 분류기를 이용한 분류를 진행한다. [Fig. 5]는 본 논문에서 제안하는 컨볼루션 레이어의 구조를 나타낸다. 그림을 보면, 전체적인 네트워크 구조는 총 3번의 컨볼루션 연산과 풀링(pooling)연산으로 구성되고 각각의 컨볼루션 연산 이후에는 활성화 함수인 ReLU를 거치고 나서 풀링 연산을 수행한다. 여기서 컨볼루션 연산은 3x3 컨볼루션 필터를 사용하며 이미지의 크기를 유지하기 위해 패딩(padding)값을 준다. 그리고 풀링의 경우 2x2 최대 풀링(max pooling)을 사용하며 이로 인해 영상의 크기가 반씩 줄어든다. 또한 마지막 풀링 연산 이후에 구조 변형(reshape)을 통해 레이어를 1차원으로 변경해주며 변형된 레이어를 완전 연결(fully connection)시켜 최종 레이어를 생성한다. 마지막으로, 최종 레이어를 분류기에 거쳐 사람인지 사람이 아닌지를 판단한다. 본 논문에서는 분류기로 softmax 함수와 kernel-SVM을 적용하여 성능을 비교 및 평가한다.

본 논문에서 제안하는 네트워크는 이전 단계에서 추출된 객체 후보자들의 크기가 작기 때문에, 입력 크기가 작은 컨볼루션

레이어 구조를 제안한다. 또한, 네트워크의 입력 사이즈는 객체의 크기 조정 중 발생하는 왜곡을 최소화하기 위해 평균적으로 학습 데이터에서 나타나는 크기인 32x64로 설정하며 해당 비율을 유지한다.

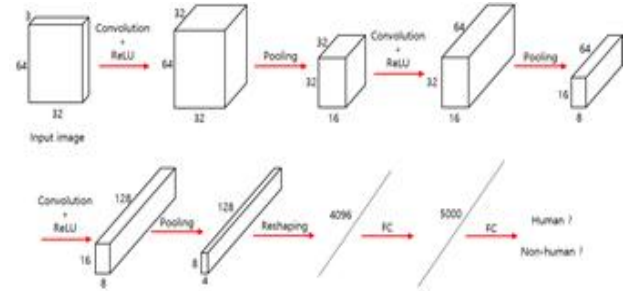


Fig. 5. The proposed network structure

네트워크에 사용된 컨볼루션 연산은 식 (6)와 같이 동작한다.

$$S(x) = (x*w)(t) \quad (6)$$

식(6)에서 x는 입력이며 w는 커널(kernel), S는 특징 맵으로, 입력과 커널의 연산을 통해 특징 맵 S를 도출한다. 이런 컨볼루션 연산이 일반적인 neural network와 다른 점은 하나의 커널이 모든 입력과 연결된 것이 아니라 희소하게(sparse) 연결된다는 점이다.

제안하는 네트워크 구조에서 각 층의 활성화 함수(activation function)는 ReLU (Rectified Linear Unit)[18]를 사용하며 식 (7)과 같이 동작한다.

$$RELU(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \quad (7)$$

식 (7)을 자세히 설명하면 다음과 같다. x는 입력 값이며 x가 0보다 작으면 출력 값이 0으로 도출되고 x가 0이거나 0보다 크면 출력 값이 x로 도출된다. 이러한 동작 과정을 통해 그레디언트(gradient)값이 너무 큰 값이나 작은 값으로 치우쳐 학습이 안되는 vanishing/exploding gradient 현상을 해결할 수 있다. 또한, ReLU는 선형함수이기 때문에 미분 계산이 간단하다는 장점이 존재한다. 따라서 본 논문에서는 여러 가지 장점을 가진 ReLU를 활성화 함수로 사용한다. 또한 학습 시에는 adam optimizer[19]를 사용한다.

마지막으로, 입수자 추적 단계에서는 객체 분류 단계에서 입수자라고 분류된 객체들에 대해 화면에 표시를 하고 KCF tracker[17]로 추적한다. 각 단계는 모두 임의로 지정된 수영 경계선을 넘어간 객체에 대해서만 진행되며 추적하던 객체가 수영경계선 안으로 들어올 경우 추적을 중단한다.

V. Experiment Results

본 논문에서는 기존 HOG 기반 입수자 검출 알고리즘과 제안하는 CNN 기반 입수자 검출 알고리즘의 성능을 비교 평가한다. 실험을 수행한 컴퓨터의 하드웨어 플랫폼은 Intel Core i5 3.3GHz, 16.0GB RAM이며, GPU는 NVIDIA GeForce GTX 1060 6GB를 사용한다. 또한, 소프트웨어 플랫폼은 Windows 10 Operate System이며 visual studio Ultimate 2013에서 opencv 3.2를 사용한다. 네트워크에 대한 학습은 FC(Fully Connected) layer를 제외한 네트워크에 대해 ImageNet[20] 데이터로 학습을 진행한다. 학습이 완료된 이후에 해당 네트워크의 가중치 값을 고정시킨 후 FC layer를 네트워크 끝에 붙여 직접 실영상에서 추출한 사람 데이터와 사람이 아닌 데이터 2,500개, 총 5,000개의 데이터로 학습을 진행한다.

실험은 1,000개의 데이터(사람 데이터 500개, 사람이 아닌 데이터 500개)를 실영상에서 추출해 진행한다. 또한, 1,136x408 크기의 영상 3개에 대해 임의의 수영경계선을 지정하고 객체를 탐지한다. 단, 분류에 대한 정확도의 테스트를 위해 tracker를 제외한 실제 프로그램으로 실험을 진행한다. CNN의 분류기로는 softmax 분류기와 rbf kernel SVM을 사용하여 실험하고 이를 비교 평가한다. 마지막으로, 실험에서 사용한 정확도는 식 (8)의 Accuracy를 사용한다.

$$Accuracy(\%) = \frac{correct\ answer}{Total\ no.\ of\ data} * 100 \quad (8)$$

식 (8)에서 분모는 테스트에 사용되는 총 데이터의 수이며 분자는 분류기가 올바르게 분류한 데이터의 수이다. 여기에 100을 곱하여 퍼센트로 나타낼 수 있고 이를 이용하여 분류기의 정확도를 측정 및 비교한다.

표 1은 테스트 데이터 1,000개에 대해 기존 HOG-SVM 방법과 본 논문에서 제안한 CNN 기반 방법(Softmax 분류기, kernel-SVM)를 사용한 경우의 정확도이다. 표 1을 보면, CNN을 사용하여 특징 벡터를 추출할 경우 HOG를 이용했을 때 보

다 약 6% 정도 높은 정확도를 보였다. 또한, 분류기로 kernel-SVM을 사용한 정확도가 softmax 분류기를 사용한 경우 보다 0.4% 높은 결과를 보였다.

Table 1. Accuracy for classifiers

	Conventional method	ConvNet-Softmax	ConvNet-SVM
Accuracy(%)	93.69	99.1	99.5

[Fig. 6]은 기존 알고리즘과 제안하는 알고리즘에 대한 결과 영상이며, [Fig. 7]은 실험에서 사용한 분류기의 성능 차이를 보여 주기 위해 실험 영상의 일부를 확대한 결과이다. 먼저, 그림 [Fig. 6-(a)]는 기존 HOG를 사용한 방법의 결과 영상으로, 앞서 언급했던 문제점들로 인해 입수자를 정확하게 검출해내지 못하는 경우가 자주 발생했다. 반면, [Fig. 6-(b)](Softmax classifier)와 [Fig. 6-(c)](kernel-SVM)를 보면, 기존 방법보다 입수자를 보다 정확하게 검출하는 것을 알 수 있다. 이와 같은 결과로 미루어 보아, 제안하는 CNN 기반 입수자 검출 알고리즘은 동적 배경에 강인하며 해안경계선을 침범한 입수자에 대해 99%이상의 정확도를 나타내는 우수한 알고리즘이라 사료된다.

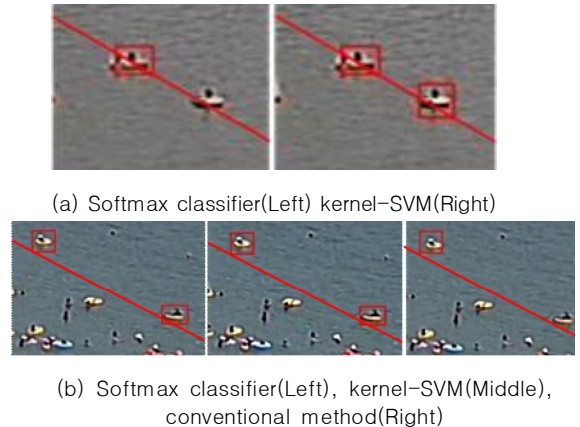
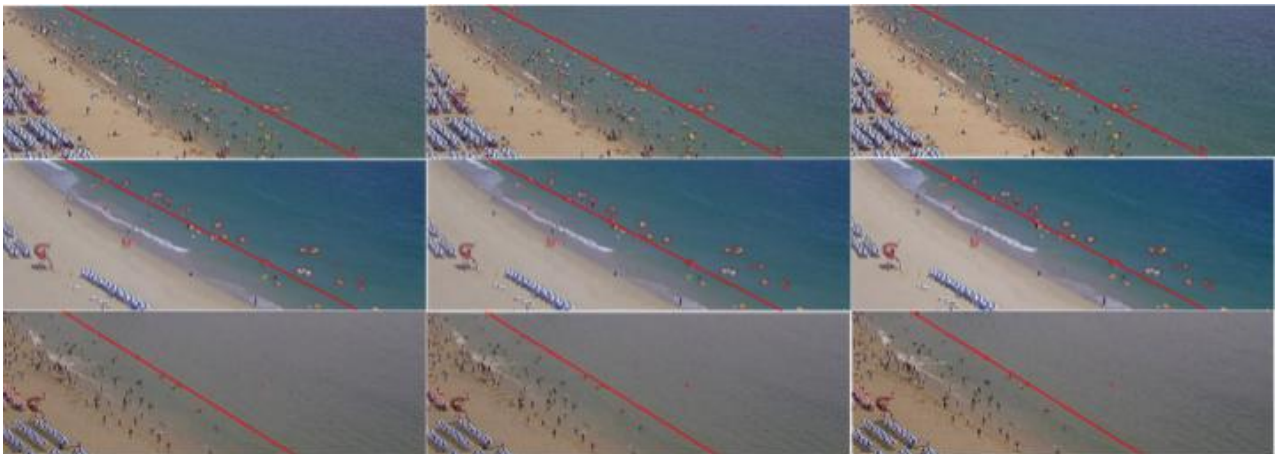


Fig. 7. The enlarged images of a part of the result image



(a) Conventional method

(b) Softmax classifier

(c) Kernel-SVM

Fig. 6. Images of experimental result using actual image data

VI. Conclusions and future work

본 논문에서는 기존 HOG 기반 입수자 검출 알고리즘에서 발생하는 긍정 오류와 부정 오류를 완화하기 위해 CNN 기반 입수자 검출 알고리즘을 제안하였다. 또한, 입수자들에 대한 분류 성능 비교를 위해 CNN의 분류기를 softmax 분류기와 kernel-SVM을 각각 사용하여 이를 비교 평가하였다. 실험 결과, 제안한 CNN 기반 입수자 검출 방법이 기존 HOG 기반 방법보다 약 6% 높은 성능을 보였다. 또한, softmax 분류기와 kernel-SVM 각각에 대해 실험한 결과, kernel-SVM을 분류기로 사용했을 때 실 영상 데이터에서 사람 객체를 좀 더 강인하게 잡는 모습을 보였다. 향후 연구로는 실영상과 일반적인 영상에서 보다 향상된 분류 성능을 위한 네트워크의 구조를 연구할 예정이다.

REFERENCES

- [1] Ministry of Public Safety and Security, "2016 Ministry of Public Safety and Security Statistics Annual Report," pp. 49-53, 2016.
- [2] YuJung Choi, Yoon Kim, "Object Detecting and Tracking Algorithm In The Environment of Complex Beach," Korean Institute of Information Scientists and Engineers Academic Publications, Vol. 2016, No. 12, pp. 1123-1125, Dec. 2016.
- [3] Z. Zivkovic, "Improved Adaptive Gaussian Mixture Model for Background Subtraction," Pattern Recognition, Vol. 2, pp. 28-31, Aug. 2004.
- [4] N. Dalal and B. Triggs, Histogram of Oriented Gradients for Human Detection, In Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition, pp. 886-893, San Diego, CA, June 2005.
- [5] C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, Vol. 2, Issue 2, pp. 121-167, June 1998.
- [6] Y. Bengio, "Learning Deep Architectures for AI," Foundations and Trends in Machine Learning, Vol. 2, No. 1, pp. 1-127, Nov. 2009.
- [7] Y. Tang, "Deep Learning using Linear Support Vector Machines," ArXiv preprint arXiv:1306.0239, June 2013.
- [8] S. Paisitkriangkrai, C. Shen, and A. Hengel, "Pedestrian Detection with Spatially Pooled Features and Structured Ensemble Learning," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 38, Issue 6, pp. 1243-1257, Aug. 2016.
- [9] H. Jiang and E. Learned-Miller, Face Detection with the Faster R-CNN, In Proc. of the Int'l Conf. on Automatic Face and Gesture Recognition, pp. 650-657, Washington, DC, May-June 2017.
- [10] JiWoong Jang, GooMan Park, "License Plate Recognition System based on Normal CCTV," Journal of the Institute of Electronics and Information Engineers, Vol. 54, No. 8, pp. 89-96, Aug. 2017.
- [11] S. Krig, "Interest Point Detector And Feature Descriptor Survey," Computer Vision Metrics, pp. 187-246, 2016.
- [12] P. Viola and M. Jones, Rapid Object Detection Using a Boosted Cascade of Simple Features, In Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition, pp. 1-9, Kauai, HI, Dec. 2001.
- [13] P. Felzenszwalb, D. McAllester, and D. Ramanan, A Discriminatively Trained, Multiscale, Deformable Part Model, In Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition, pp. 1-8, Anchorage, AK, June 2008.
- [14] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," Computer Vision and Pattern Recognition, arXiv:1612.08242, Dec. 2016.
- [15] R. Girshick, Fast R-CNN, In Proc. of the IEEE Int'l Conf. on Computer Vision, pp. 1440-1448, Santiago, Chile, Dec. 2015.
- [16] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, In Proc. of the Int'l Conf. on Advances in Neural Information Processing Systems, pp. 91-99, Quebec, Canada, Dec. 2015.
- [17] F. J. Huang and Y. LeCun, Large-Scale Learning with SVM and Convolutional for Generic Object Categorization, In Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition, pp. 284-291, New York, NY, June 2006.
- [18] J. F. Henriques and R. Caserio, "High-Speed Tracking with Kernelized Correlation Filters," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 37, Issue 3, pp. 583-596, Aug. 2015.
- [19] V. Nair and G. E. Hinton, Rectified Linear Units Improve Restricted Boltzmann Machines, In Proc. of the 27th Int'l Conf. on Machine Learning, pp. 807-814, Haifa, Israel, June 2010.
- [20] K. Diederik and J. Ba, "Adam: A Method for Stochastic Optimization," ArXiv preprint arXiv:1412.6980, Dec. 2014.
- [21] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei., ImageNet: A Large-Scale Hierarchical Image Database, In Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition, pp. 248-255, Miami, FL, June 2009.

Authors



Dasol Hong received the B.S. degrees in Computer Science from Kangwon National University, Korea, in 2016. Hong joined the Department of Computer Communication and Engineering at Kangwon National University, Kangwon, Korea, in 2016. She

is currently a student in the Department of Computer Communication and Engineering, Kangwon National University. She is interested in image processing, computer vision and machine learning.



Yoon Kim received the B.S., M.S. and Ph.D. degrees in Electronics Engineering from Korea University, Korea, in 1993, 1995 and 2003, respectively. Dr. Kim joined the faculty of the Department of Computer Communication and Engineering at

Kangwon National University, Kangwon, Korea, in 2004. He is currently a Professor in the Department of Computer Communication and Engineering, Kangwon National University. He is interested in computer vision, image processing, machine learning, digital signal processing.