

인공지능 이미지 인식 기술을 활용한 위험 알림 CCTV 서비스

이하린*, 김유진*, 이민아*, 문재현***

*홍익대학교 컴퓨터공학과

**한국기술거래사회

+교신저자

rinchive@gmail.com, dbwls3225@naver.com, hdmksm1@naver.com,

smjhoon@gmail.com

Danger Alert Surveillance Camera Service using AI Image Recognition technology

Ha-Rin Lee*, Yoo-Jin Kim*, Min-Ah Lee*, Jae-Hyun Moon***

*Dept. of Computer Science, Hongik University

**Korea Technology Transfer Agents Association

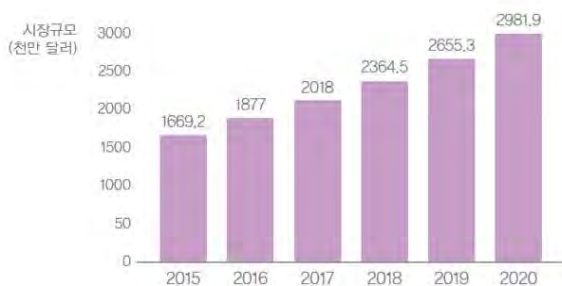
+Corresponding author

Abstract

The number of single-person households is increasing every year, and there are also high concerns about the crime and safety of single-person households. In particular, crimes targeting women are increasing. Although home surveillance camera applications, which are mostly used by single-person households, only provide intrusion detection functions, this service utilizes AI image recognition technologies such as face recognition and object detection to provide theft, violence, stranger and intrusion detection. Users can receive security-related notifications, relieve their anxiety, and prevent crimes through this service.

1. 서론

인공지능 시장규모는 폭발적인 성장세를 보이며 인공지능과 융합된 산업들의 영역 또한 넓어지고 있다. CCTV 역시 아날로그에서 지능형 CCTV로 대체가 넘어가며 단순한 감시 기능을 넘어 딥러닝 기반의 영상 인식 기술을 더해 더욱 고도화된 기능을 갖게 되었다. [그림 1]에 따르면 전 세계 지능형 CCTV 시장은 2015년 166억만 달러에서 2020년에는 298억만 달러의 대규모 시장을 형성할 것으로 예측된다. 또한 국내에서도 과학기술정보통신부는 2019년 '5G+ 전략'을 발표했고, 5G를 매개로 한 10대 핵심 산업 중 하나로 지능형 CCTV를 선정했다. 따라서 지능형 CCTV는 영상보안의 핵심으로 자리 잡아 널리 이용되는 날이 머지않을 것으로 보인다.



(그림 1) 전 세계 인공지능 CCTV 시장성장 추이[1]

2. 문제 제기

기존 CCTV는 실시간으로 이상 현상을 감지하기 위해 관제요원의 모니터링이 필요하다. 하지만 주거 방범용 CCTV의 관제요원 배치는 쉽지 않을뿐더러 CCTV 관제요원이 감시 시작 후 10분이 지나면 위험 상황의 50%, 22분이 지나면 95%를 놓쳐 범죄 예방에 큰 어려움이 있다.[2] 따라서 CCTV가 범죄의 사전적 예방 역할을 하기 위해서는 이상 현상 감지 시 사용자에게 즉각적인 알림을 주어야 한다. 본 서비스는 AI CCTV로서 배회자, 도난, 위협 행위, 침입자, CCTV 무력화를 감지하고 사용자에게 즉각적인 푸쉬 알림을 전송하여 기존의 문제점을 보완했다.

3. 문제 범위

본 서비스는 보안이 취약한 여성 1인 가구를 타겟으로 삼는다. 2018년 인구주택총조사에 따르면 2000년 이후 1인 가구 증가율은 전체 일반 가구 증가율의 5배 가까이 늘어났다.[3] 일반 가구 대비 1인 가구의 비율은 2000년 15.5%에서 연속적으로 증가해 2017년에 28.6%를 차지했다. 또한 1인 가구가 급증하면서 1인 가구의 범죄 취약성이 사회 문제로 떠오르고 있다. 2012년 국민생활안전실태조사 '주거침입 관련 범죄 피해가구 특성'에 관한 연구 결과에 따르면, 1인 가구를 대상으로 한 범죄는 전체 범죄 건수의 30% 이상을 차지한다. (최상희 외, 2013)[4]



(그림 2) 사회 안전에 대한 인식[5]

또한 [그림 2]에서 보듯 2016년 여성 1인 가구의 46.2%가 사회의 전반적인 안전에 대해 불안하다고 답했으며 안전하다고 느끼는 비율은 13%에 그쳤다. 따라서 본 서비스는 1인 가구 사용자가 이상 현상 감지에 대한 즉각적인 알림을 받으며 불안감을 해소하고 범죄를 예방하는 데 집중한다.

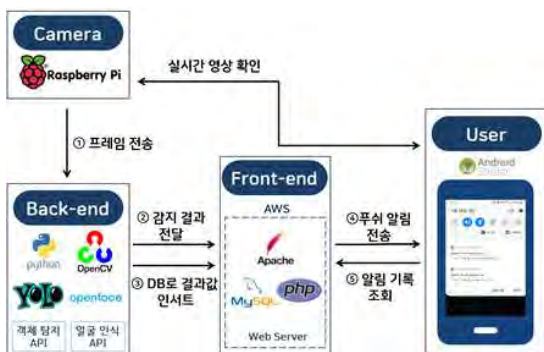
4. AI CCTV 서비스

4.1 AI CCTV 서비스 개요

본 서비스는 사용자가 거주지의 위험 상황에 대해 어플리케이션을 통해 실시간으로 알림을 받을 수 있다. 거주지에서 일어날 수 있는 도난, 위협 행위, 외부인, 침입자, 카메라 무력화를 위험으로 간주한다. 인공지능 객체 탐지 알고리즘과 얼굴 인식 알고리즘을 활용하여 외부인, 도난, 위협 행위를 감지할 수 있도록 한다. 또한 AWS(Amazon Web Server)를 이용해 백엔드 서버와 카메라와 연결된 라즈베리파이로부터 정보를 전달받아 데이터베이스를 관리하고 사용자에게 알림을 전송한다.

4.2 AI CCTV 서비스 구성도

본 서비스는 라즈베리파이, 딥러닝 연산을 위한 백엔드 서버, 웹서버, 안드로이드 어플리케이션으로 구성되어 있다. 원활한 인공지능 알고리즘 연산을 위해 백엔드 서버 컴퓨터에 'GTX 1660 SUPER' 그래픽카드를 사용했고, AWS에서 Apache, MySQL, PHP로 웹서버를 구축했다.



(그림 3) 서비스 구성도

[그림 3]은 본 서비스의 구성도이다. 라즈베리파이에 연결된 카메라를 통해 얻은 영상을 실시간으로 백엔드 서버로 전송하고, 백엔드 서버는 인공지능 객체 탐지 알고리즘

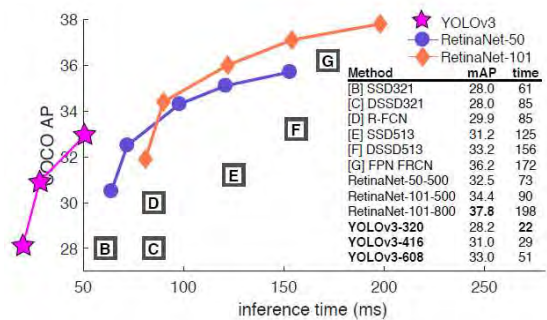
'YOLOv3'와 얼굴 인식 알고리즘 'OpenFace'를 이용하여 영상 속 위험 상황을 감지한다. 위험 상황이라고 판단된 경우 웹서버로 감지 결과를 전송하고 웹서버에서 안드로이드 어플리케이션으로 푸시 알림을 전송한다. 알림을 전송 받은 사용자는 어플리케이션을 통해 알림에 대한 세부 정보를 확인할 수 있고 실시간으로 카메라 영상을 확인할 수 있다.

또한 사용자는 어플리케이션을 통해 경비 모드를 설정할 수 있다. 경비 모드 작동 시 현관에 설치된 모션 센서를 통해 움직임이 감지될 경우 사용자가 경비 모드를 해제할 때까지 1분마다 반복적으로 푸시 알림을 사용자에게 전송하여 침입자 발생을 알린다.

5. 인공지능 알고리즘 선정 및 개선

5.1 객체 탐지 알고리즘 선정

객체를 탐지하기 위한 대표적인 알고리즘으로 CNN과 YOLOv3가 알려져 있다. CNN 알고리즘은 제안된 영역이 어떤 클래스에 속하는지 분류하기 위해 한 이미지에 수천 개의 네트워크를 거쳐 정확도는 높지만 속도가 느린 것이 특징이다. 본 서비스에서 도난, 위협 행위, 배회자를 실시간으로 감지하기 위해서 빠르고 정확한 결과를 얻을 수 있는 YOLOv3 Object Detection 알고리즘을 사용한다. YOLOv3는 한 프레임에 대해서 클래스 분류와 Bounding Box 찾기가 동시에 이루어지고, 인풋 이미지 전체를 한 번 본 후 이미지의 global context를 바탕으로 예측이 가능하여 inference time이 매우 짧다. [그림 4]는 다른 Detection model들과 YOLOv3의 성능을 비교한 결과이다. YOLOv3는 단 하나의 네트워크를 통해 결과값을 얻을 수 있어 R-CNN에 비해 1000배, Fast R-CNN에 비해 100배가량 빠른 성능을 보인다.



(그림 4) 다른 Detection Model과 성능 비교[6]

5.2 YOLOv3 예측률 개선

위험 행위와 도난을 감지하기 위해 본 서비스는 YOLOv3 객체 탐지 알고리즘을 사용한다. 흥기를 들거나, 뒤에서 제압하는 모습을 위협 행위로 간주했고, 객체 탐지 모델 학습을 위해 해당하는 동작을 실제로 재연한 사진을 촬영했다. 도난 감지 물품으로 택배, 우산, 자전거를 선정하여 일반적으로 사용되는 다양한 크기의 택배 상자와 택배 봉투

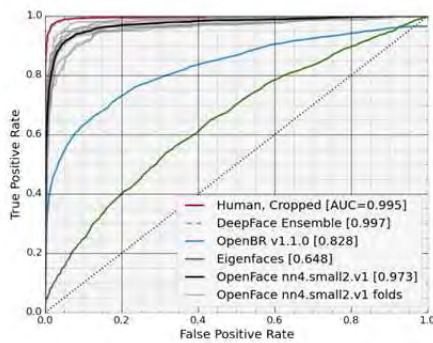
를 촬영했고, 우산과 자전거 이미지 데이터는 COCO 데이터 셋[7]을 사용했다. 모델 학습을 위해 클래스별로 이미지를 2000장씩 수집했고, 객체의 일반적인 이미지와 CCTV 카메라가 포착하는 객체의 모습이 다른 것을 고려하여 학습 이미지의 각도를 조절함으로써 예측률을 개선했다. [표 1]은 수집한 이미지로 학습하여 얻은 분류기 모델의 예측률을 나타낸다. CCTV 각도를 고려하지 않은 이미지 데이터를 사용했을 때는 평균 50%의 예측률을 보였지만 각도와 위치를 고려했을 때는 약 75%의 개선된 예측률을 보였다.

(표 1) 학습 이미지에 따른 객체 예측률

	CCTV 각도를 고려하지 않은 이미지	CCTV 각도를 고려한 이미지
택배상자	0.64	0.86
택배봉투	0.5	0.64
흥기를 든 행위	0.6	0.9
뒤에서 제압하는 행위	0.26	0.58

5.3 얼굴 인식 알고리즘 선정

본 서비스는 YOLOv3를 통해 사람 객체를 감지한 후, 얼굴 인식을 통해 낯선 사람이 배회하는지 판단해야 한다. OpenFace, DeepFace, OpenBR 등 여러 얼굴 인식 알고리즘이 있지만, 그 중 OpenFace는 분류기 모델 훈련 속도가 준수하며 python API를 제공한다. OpenFace는 학습된 DNN 모델을 사용하여 각 얼굴에서 128개의 특징점을 뽑아 얼굴을 분류한다.



(그림 5) ROC곡선의 AUC 수치 비교[8]

[그림 5]는 머신러닝 모델의 평가 지표인 ROC 곡선의 AUC 수치를 비교한 자료이다. Principle Component Analysis를 활용한 Eigenface 모델의 경우 0.648, Spectrally Sampled Structural Subspaces Features 알고리즘의 OpenBR 모델은 0.828의 결과를 갖는다. 이에 비하여 OpenFace nn4.small2.v1 모델은 0.973의 AUC로 높은 성능을 보였다. 따라서 본 서비스는 python API를 제공하고 정확도가 우수한 OpenFace를 얼굴 인식 알고리즘으로 채택하여 사용한다.

5.4 OpenFace 오탐 및 미탐 개선

OpenFace를 통해 생성된 분류기 모델은 학습된 사람 클래스 외의 다른 사람의 얼굴 이미지가 주어지면 학습된 클래스 중에서 결과를 내는 문제가 발생한다. 즉 A와 B의 얼굴을 학습시킨 모델에 C의 이미지를 인풋으로 주었을 때, 학습되지 않았음을 결과로 내지 않고 A와 B 중 가장 비슷한 인물로 결과를 예측하는 것이다. 본 서비스에서는 사용자가 등록하지 않은 사람을 외부인으로 구분하기 위해 'Unknown'이라는 외부인 클래스를 추가하여 예측 결과로 Unknown을 낼 수 있도록 했다. 사용자는 어플리케이션을 통해 얼굴이 잘 보이는 사진을 선별해 인물별로 10장씩 DB에 등록한다. 사용자가 등록한 사진을 안전한 인물 클래스의 학습 데이터로, AI 허브(aihub.co.kr)에서 제공하는 한국인 안면 인식 데이터를 Unknown 클래스 학습 데이터로 사용해 SVM(Support Vector Machine) 지도 학습 모델인 얼굴 인식 분류기 모델을 생성한다.

얼굴 인식 분류기는 인풋 이미지에서 포착한 얼굴의 특징점을 수치화한 128차원의 임베딩 값과 각 클래스의 임베딩 값을 비교하여 가장 가까운 임베딩 값을 가지는 클래스를 결과로 예측한다. 본 서비스는 Unknown 클래스 학습에 다양한 연령대의 남녀 사진을 여러 장 사용하여, 모르는 사람의 얼굴을 인식하는 과정에서 등록된 사람 클래스의 임베딩 값보다 Unknown 클래스의 임베딩 값과 가까울 수 있도록 했다.

분류기 모델이 외부인을 등록된 사용자로 구분하는 미탐(false negative)을 해결하기 위해 Unknown 클래스 학습 데이터 수를 조정했다. 얼굴 인식 분류기 모델은 지도 학습을 통해 생성되기 때문에 각 클래스의 데이터가 늘어날수록 정확한 답을 예측한다. 따라서 등록된 인물 외의 다양한 인물 이미지를 Unknown 클래스 학습에 더 많이 사용할수록 외부인을 Unknown 클래스로 예측할 확률을 높일 수 있다.

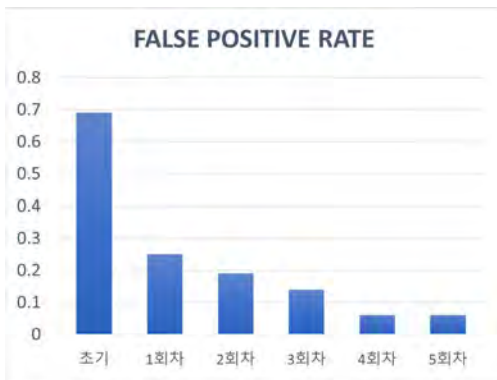
(표 2) Unknown 이미지 데이터 수에 따른 오탐률과 미탐률

	Unknown 클래스의 데이터 수		
	100장	300장	500장
False positive rate	0.14	0.24	0.45
False negative rate	0.27	0.07	0.05

학습을 위한 최적의 데이터 수를 찾기 위해 Unknown 클래스의 데이터 수를 달리하여 모델을 생성했다. [표 2]는 Unknown 클래스의 학습 데이터 수에 따른 테스트 결과이다. 100장은 27%, 300장은 7%, 500장에 대해서 5%의 미탐률을 보였다. 500장을 사용했을 때 가장 낮은 미탐률을 보였지만, 300장과 500장의 미탐률 차이가 근소하여 더 빠른 모델 생성을 위해 Unknown 클래스 학습에 300장의 이미지를 사용했다. 다만 Unknown 클래스의 데이터를 늘려 미탐을 해결할수록 오탐률(false positive rate)이 증가하게

된다. [표 2]에서도 Unknown 클래스의 데이터 수가 증가할수록 오답률이 증가하는 모습을 보였다.

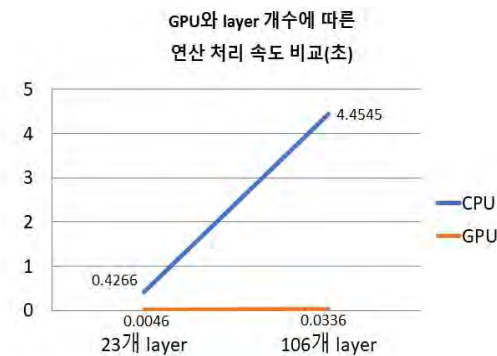
Unknown 클래스 데이터 수 증가에 따른 오답률을 개선하기 위해 Unknown 클래스로 잘못 예측된 사진을 해당 클래스의 학습 데이터 셋에 새로 추가하여 모델을 다시 학습시켰다. 만약 사용자가 잘못된 결과에 대해 알림을 받았을 때, 사용자는 어플리케이션을 통해 서버와 DB로 잘못된 결과가 전송되었음을 알릴 수 있다. 이후 서버는 DB에서 잘못 예측된 데이터를 가져와 해당하는 클래스에 데이터를 추가하고 새로운 분류기 모델을 생성해 오답을 줄일 수 있도록 했다.



(그림 6) 재학습 횟수에 따른 오답률

[그림 6]은 Known 클래스에 10장, Unknown 클래스에 300장을 사용해 학습한 모델의 재학습 횟수에 따른 오답률을 나타낸다. 테스트 셋 중에서 오답의 결과를 낸 사진을 한 장씩 추가하여 다시 학습한 결과, 오답률이 확연히 줄어드는 모습을 보였다. 따라서 본 서비스는 사용 초반 분류기 모델의 재학습 과정을 통해 얼굴 인식 오답률을 줄였다.

6. 인공지능 학습 및 객체 탐지 속도 개선



(그림 7) GPU와 layer 개수에 따른 연산 속도

YOLOv3에서 기본으로 제공하는 신경망은 106개의 레이어로 이루어져 있다. 본 서비스에서는 기존 106개 층의 신경망을 23층으로 줄여 처리 속도를 높이는 것을 시도했다. 이미지 200장으로 테스트한 결과, 106개의 신경망을 사용했을 때는 약 70%의 정답률을 보인 반면 23개의 신경망의 경우 약 12%의 정답률을 보였다. 따라서 신경망 레이어

수를 줄일 경우 예측 속도는 개선할 수 있으나 정확도는 떨어지게 된다. 정확도 유지를 위하여 레이어를 줄이는 방법은 사용하지 않고, 병렬 연산을 위한 GPU를 추가하여 속도를 향상하도록 했다. GPU 사용을 위해 CUDNN과 CUDA 라이브러리를 설치하여 백엔드 서버에서 GPU를 사용할 수 있도록 했다. [그림 7]은 GPU 여부와 레이어 개수에 따른 연산 속도를 보여준다. 같은 이미지에 대한 객체 탐지 결과, GPU를 사용했을 때 약 133배 빨라져 정확도를 떨어뜨리지 않고 속도를 개선했다. 또한 CPU만으로 객체 탐지 모델을 학습할 경우 약 한 달의 시간이 소요되지만, GPU를 사용하면 4일 이내에 학습이 끝나기 때문에 약 8배 빠른 학습이 가능하다.

7. 결론

본 서비스는 객체 탐지 알고리즘과 얼굴 인식 알고리즘을 이용한 지능형 CCTV 서비스에 관해 기술하였다. 특히 두 알고리즘의 오답과 미탐을 줄여 서비스의 신뢰성을 높였다. 또한 서비스의 속도를 향상할 수 있는 방안 중점을 두어 GPU를 이용해 인공지능 모델 학습 및 객체 탐지를 더욱 빠르게 처리할 수 있게 했다. 푸쉬 알림을 통해 사용자가 CCTV를 확인해야 하는 불편을 줄였고, 딥러닝을 이용한 이미지 인식 기술을 활용하여 사용자에게 자세한 알림 서비스를 제공하여 거주지의 안전성을 더욱 강화할 수 있다. 이러한 이미지 인식 기술을 갖춘 지능형 CCTV 알림 서비스는 출결 확인이나 출입 관리 등 얼굴 인식이 필요한 다른 분야에서도 활용될 수 있다.

참고문헌

- [1] 시장조사기관 IMS리서치 자료종합(2019)
- [2] Fleck S., Straßer W, "Privacy Sensitive Surveillance for Assisted Living - A Smart Camera Approach. Handbook of Ambient Intelligence and Smart Environments", pp.985-1014, 2010.
- [3] 통계청, 인구주택총조사에 나타난 1인가구의 현황 및 특성 보도자료, 2018.
- [4] 최상희·정소이·김용태·정경석, 「도시형생활주택의 평가 및 발전방향 연구」 한국토지주택공사토지주택연구원, 2013.
- [5] 박병률, "여성 1인 가구 46% '사는 게 불안해요'", 경향신문, 2017년 06월 27일자.
- [6] Joseph Redmon and Ali Farhadi, "YOLOv3: An Incremental Improvement", pp. 1, 2018.
- [7] Lin, Tsung-Yi, et al. "Microsoft COCO: Common Object in Context", ECCV, 2014.
- [8] Brandon Amos, Bartosz Ludwiczuk, & M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications", pp. 9, 2016.

본 논문은 과학기술정보통신부
정보통신창의인재양성사업의 지원을 통해 수행한
ICT멘토링 프로젝트 결과물입니다.