딥러닝을 이용한 범죄예방 현금인출기

박지우, 이하영, 백화영, 박보영, 조중휘
인천대학교 임베디드시스템공학과
email : {inu201401651,leehy0321,hwayoung,inu201401646,jcho}@inu.ac.kr

ATM for criminal prevention using deep learning

Jiwoo Park, Hayoung Lee, Hwayoung Baek, Boyoung Park, Junghwi Jo
Dept. of Embedded System Engineering, Incheon University

요 약

본 논문은 보이스피싱 발생 후의 대처 방안이 아닌, 범죄 행위 자체의 예방을 목적으로 하는 신원 허가 후의 자동인출기 시스템을 제안한다. 범죄예방 현금인출기(ATM)의 작동과정은 크게 두 가지이다. 첫째, YOLO Detection System을 이용하여 학습된 데이터에 기반하여, 사용자의 얼굴에서 선글라스와 마스크를 검출한다. 둘째, 미리 학습된 범죄자 모델 데이터에 앞서 사용자의 신원을 조회하고 ATM의 사용허가를 내준다. 혹은 주요지명 피의자일 경우, 경찰에 실시간 안내를 주어 범죄 수사를 용이하게 한다.

1. 서론

2017년 금융감독원의 발표에 의하면, 한국의 보이스피싱 피해자는 5만 명, 보이스피싱 피해액은 2천400억에 달한다. 그러나 예방된 사건은 2016년에 17건, 2017년에 40건 정도밖에 되지 않는다. 일반적인 보이스피싱 범죄의 경우, 범죄자가 피해자와 전화 통화를 한 후, ATM에서 현금을 인출하는 방식으로 범행을 진행하는데 [1]의 논문에 따르면 한국의 ATM 거래와 같은 비대면 거래형태는 대면거래보다 약 8배 높은 수치를 나타내며 이를 통해 우리나라가 보이스피싱 범죄에 많은 노출이 되어있음을 알 수 있다. 현재 이러한 보이스피싱의 검거율이 낮은 이유는 보이스피싱 현금 인출책들이 모자를 깊이 눌러쓰고 마스크를 쓴 모습인 경우가 많아 CCTV를 통해서는 ATM 사용자의 신원 확인이 어렵기 때문이다. CCTV와 ATM기에 부착되어 있는 카메라가 범죄자의 신원 확인을 위해 사용되고 있기는 하나 범죄 근절에 활용적으로는 도움이 되지 못하고 있는 상황이다. 최근에는 사용자의 지문이나 홍채 등의 신체특성을 감지하여 신원을 판별하는 기술들이 개발되어 간단하게 사용자의 신원을 판단할 수 있고 인식률 또한 높지만 기존의 ATM에 추가적으로 장비를 설치해야 하기 때문에 많은 비용을 필요로 한다는 문제가 있다.[2]

본 연구는 기존의 ATM을 이용하여 범죄 예방과 동시에 범인 검거에 도움을 주는 '범죄 예방 ATM''을 제안한다. '범죄 예방 ATM'은 영상처리, 얼굴인식, 딥러닝(Deep-Learning) 기술을 활용하여 ATM 사용자의 드러낸 얼굴을 촬영해 서버의 DB에 저장하며 사용자가 범죄자일 경우, 관리자에게 범죄자의 정보를 전달하고 촬영된 ATM 사용자들의 얼굴은 범죄자 검거를 위해 조회할 수 있도록 설계하였다.

* 이 논문은 2018 한이음 ICT 멘토링 프로젝트의 연구비 지원을 받은 결과물입니다.

2. 구성도 흐름

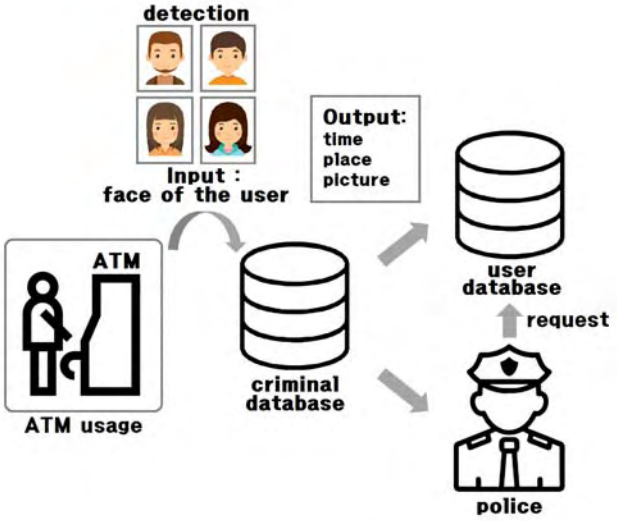

그림 1. 시스템 구성도

그림 1은 '범죄 예방 ATM'의 시스템 구성도이다. ATM에 접근한 사용자가 있는지 PIR 센서를 이용하여 판단한 후. CDS 센서를 이용하여 사용자의 얼굴을 촬영하기 위한 적정한 밝기로 램프를 킨다.

촬영하기 위한 조건이 만족되었다면 사진 캡처를 위한 얼굴인식(Face Recognition)을 진행한다. 얼굴이 정상적으로 식별된다면 딥러닝(Deep Learning) 기술로 학습시킨

선글라스 혹은 마스크가 검출된다면 사용자에게 액세서리 탈의를 요청하는 음성 및 텍스트 메시지를 출력한다.

정상적으로 드러난 사용자의 얼굴이 촬영되었다면 이것을 내장된 범죄자 DB와 비교하여 현재 사용자가 범죄자인지 아닌지를 판단한다. 범죄자라고 판단될 경우, 담당 관리자에게 범죄자의 신원과 촬영된 사진 및 위치, 시간 정보를 보내 범인 검거에 도움을 주도록 한다. 이후 촬영된 사진을 서버 DB에 저장하여 이후 범죄 예방 및 검거에 사용한다.

3. 시스템 구현

3.1 구현환경

구분		항목	적용내역
S/W 개발 환경	OS	Linux Ubuntu 16.04	Linux 환경에서 개발
	개발환경 (IDE)	PyQt, gedit, Python IDLE	PyQt로 LCD화면 ,gedit에서 C++로 구현 Python IDLE는 데이터셋을 구성할 때 사용
	개발 도구	OpenCV, Maria DB	OpenCV는 영상처리에서 필요한 함수를 이용, Maria DB는 범죄자의 이름, 사진, 주소, 나이, 신장, 체격 등의 정보가 들어있는 데이터베이스를 구축하는데 사용
	개발 언어	C, C++	영상처리 및 딥러닝을 구현하는데 사용
H/W 구성 장비	입력단자	Jetson Tx1	데이터를 학습하고 데이터를 전송에 이용
		PIR 센서, CdS 센서	사용자가 있는지 판단하고 사용자 주변의 밝기를 판단
		카메라	사용자의 얼굴을 촬영할 때 사용
	출력단자	LCD	얼굴 촬영 전 후의 알림을 LCD에 출력
		스피커	알림을 스피커를 통해 출력

그림 2 시스템의 SW/HW 환경

3.2 시스템 알고리즘

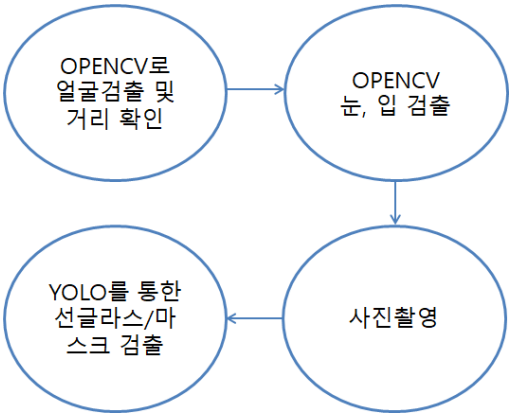

그림 3 시스템 알고리즘

사용자가 ATM에 접근하면 opencv로 얼굴을 검출하고 거리를 확인한다. 얼굴이 검출되면 opencv를 통해 사용자의 눈과 입을 검출한다. 사용자의 눈과 입이 모두 검출되면 사진을 촬영한다. 촬영된 사진에 yolo 딥러닝(Deep learning) 기술을 적용하여 선글라스 및 마스크 검출한다. 아래는 시스템의 주요 기술인 딥러닝 기술의 데이터 수집 및 학습방법에 대한 설명이다.

3.3. 데이터 수집

3.3.1 구글 크롤링 이용

사용자의 신원파악을 위해 얼굴에 장애물이 있는지를 판단하기 위한 마스크와 선글라스 사진을 구글 검색 엔진에서 크롤링하여 관련 이미지를 수집했다. 또한 안경과 선글라스를 구분해야하기 때문에 안경 이미지도 수집했다.

마스크 500개, 안경 400개, 선글라스 530개를 수집.

3.3.2 직접 수집

범죄자 사진을 직접 얻기 어려워 우리를 범죄자로 가정한 후, 카메라에서 얼굴을 자동으로 인식하여 얼굴이 담긴 이미지를 수집했다. 이미지 수집은 한명 당 1시간 30분이 걸렸으며, 클래스 당 1500개를 수집하였다.

3.4 학습모델 구축

딥러닝 기술 중에서 실시간 처리가 가능한 You Only Look Once(YOLO)를 사용한다. YOLO를 실행시키기 위해서 Darknet framework를 사용하며 Darknet은 신경망 프레임 워크로 dnn(deep neural network)를 학습시키고 실행시킬 수 있는 틀을 말한다.[3]

YOLO는 각 이미지를 S x S 개의 그리드로 분할하고, 그리드의 신뢰도를 계산한다. 신뢰도는 그리드 내 객체 인식 시 정확성을 반영한다. 다음 그림과 같이 처음에는 객체 인식과는 동떨어진 경계 상자가 설정되지만, 신뢰도를 계산하여 경계 상자의 위치를 조정함으로써, 가장 높은 객체 인식 정확성을 가지는 경계 상자를 얻을 수 있다.

다음은 YOLO의 동작 방법을 나타낸다.

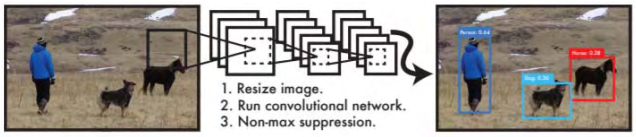

그림 4. The YOLO Detection System.[3]

YOLO는 빠르고 정확하기 때문에 실시간이 중요한 본 연구에 적합하다 판단하였고 카메라를 연결하여 실시간으로 감지되는지 검증한 후 학습모델을 구축하였다.

3.5 데이터 학습

YOLO는 이미지에서 객체를 추출하기 전에 객체의 범위에 있는 경계 상자를 먼저 확인한다. 경계 상자를 식별하기 위해서는 Class의 이름, 4변에 대한 정보 이렇게 5가지의 변수를 사용한다. 경계 상자를 예측하기 위한 그리드에 객체 포함 여부를 계산하기 위해, 객체 클래스 점수를 계산한다. 이 결과로 총 S x S x N 객체가 예측된다. 이 그리드의 대부분은 낮은 신뢰도를 가진다. 신뢰도를 높이기 위해 주변의 그리드를 합칠 수 있다. 이후, 임계값을 설정해 불필요한 부분은 제거할 수 있다.

인식 정확도를 높이기 위해 학습 데이터, 평가 데이터의 개수와 구축된 학습모델을 수정을 하며 최적의 학습모델

을 찾는 과정을 진행한다.

3.4.1 선글라스, 마스크 이미지 학습

선글라스, 마스크를 500개의 이미지를 학습 모델에 학습시킨다. 선글라스를 구별하기 위해 추가적인 데이터로 안경을 추가하였기 때문에 안경은 400개의 이미지만을 학습시킨다.

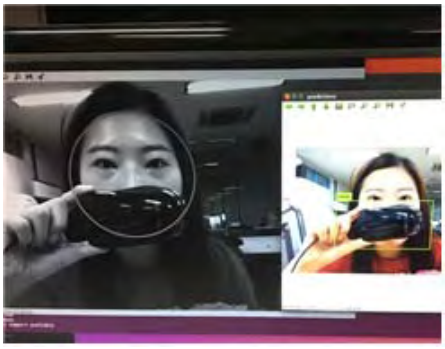

그림 5. 얼굴과 마스크 검출

3.4.2 범죄자 이미지 학습

범죄자 클래스당 1500개의 이미지(총 4명)를 학습 모델에 학습시킨다.

그림 6. 범죄자(예시) 얼굴 검출

4. 실험 결과

학습 데이터에 따른 평균 정확도에 대한 실험을 두 차례를 거쳐 진행하였다.

첫째, 학습시킨 이미지의 개수에 따른 평균 정확도

둘째, 사물 자체를 학습시켰을 때와 사물을 착용한 사람을 학습시켰을 때의 정확도의 차이

4.1 트레이닝 시킨 이미지 개수와 평균 정확도

이미지의 수(장)	10	100	300	500
평균 정확도(%)	16	32	58	63

표 2. 학습시킨 이미지의 개수에 따른 평균 정확도 결과

이미지의 개수가 증가함에 따라 평균 정확도도 높아지는 비례 관계임을 확인할 수 있다. 300장의 이미지를 학습시켰을 때와 500장의 이미지를 학습시켰을 때의 평균 정확도의 증가율이 0.025(%/1장)인 것을 고려하였을 때, 300장 이상의 이미지 학습은 큰 영향을 끼치지 않는 것으로 판단된다.

4.2 사물 자체를 학습시켰을 때와 사물을 착용한 사람을 학습시켰을 때의 정확도 차이(500장)

이미지의 내용	물체만	사람과 물체
평균 정확도(%)	58	78

표 3. 사물 자체를 학습시켰을 때와 사물을 착용한 사람을 학습시켰을 때의 정확도 차이 결과

학습시킨 이미지의 내용에 따라 평균 정확도가 달라짐을 알 수 있다. 사물 자체를 학습시켰을 때는 평균 58%의 정확도를 보이고, 사물을 착용한 사람을 학습시켰을 때는 평균 78%의 정확도를 보여준다. 이를 고려하였을 때, '사물을 착용한 사람' 이미지를 트레이닝 시키는 것이 정확한 결과를 얻을 수 있음을 확인할 수 있다.

5. 결론 및 향후 연구

범죄예방 ATM은 평균 78%의 정확도로 선글라스와 마스크를 검출하고, 평균 90%의 정확도로 얼굴을 인식하여 신원 확인에 높은 판별률을 보여주었다. 실제 ATM에 적용이 된다면, 5만명의 피해자, 2천 400억에 달하는 피해액을 대폭 줄이고, 신속한 범죄 검거가 가능할 것으로 예상된다.

또한 본 시스템은 생체인증을 시행하는 국내 은행에 안면 인식을 통한 ATM 사용 허가 시스템으로써 도입이 가능하다. 현재 얼굴의 일부를 가리는 선글라스와 마스크를 검출하고, 사용자 인식을 진행하는 방향으로 연구가 완료되었으나 향후 눈 깜빡임, 동공 움직임을 면밀히 연구하여 보안면에서 우수한 인증 방식으로 추가할 것을 기대해본다.

참고문헌

[1] 최관, 김민지 (2015). 한국 보이스피싱 범죄의 진행과정에 관한 연구. 경찰학연구, 15(3), 233-261.

[2] 홍정오, "생체인식 기술을 이용한 은행인증시스템 설계에 관한 연구", 한양대산업대학원석사논문, 2000년 6월.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Proceeding of Computer Vision and Pattern Recognition, pp.779-788, 2016*