지능형 CCTV를 위한 딥러닝 기반 안면 비식별화

Deep Learning based Face De-identification for Intelligent CCTV

박계원
Kye-Won Park

pkw3136@naver.com

동국대학교 컴퓨터공학과
Dept. of Computer Science and Engineering, Dongguk Univ., Seoul

요 약

본 논문은 지능형 감시 시스템을 구축하기 위해, YOLOv5 객체 인식 딥러닝 모델을 통해 CCTV상에서 촬영되는 영상 속의 안면 객체를 실시간으로 검출한 후 해당 객체의 영역에 대해 원본 이미지를 알아볼 수 없도록 비식별화를 하는 방법을 제시한다. 검출하는 안면 객체는 현재 마스크 착용이 강제되는 국/내외 상황을 고려하여 마스크를 착용한 얼굴을 포함시킨다. 검출 영역은 랜덤 픽셀값들로 이루어진 노이즈 파일과의 XOR연산을 통해 비식별화를 진행하여, 복호화에 필요한 원본 영상 이미지의 얼굴영역을 따로 저장하는 일이 없도록 한다. 비식별화가 완료된 영상은 수사인력이 필요시 XOR 연산을 다시 행하여 원본 영상으로 복호화를 할 수 있도록 한다. 실험은 다수의 사람 또는 빠르게 움직이는 사람이 있는 영상으로 진행하였다. 결과는 육안으로 식별이 불가능한 정도로 작은 안면 영역이나, 어두운 영역을 제외하고는 대부분의 상황에서 얼굴 검출 및 비식별화가 잘 이루어졌다. 비식별화 속도는 실시간에 쓰일 수 있는 약 25fps로 측정되었다.

키워드 : XOR연산, 노이즈 파일, 마스크 착용 얼굴 검출, 안면 비식별화, 지능형 감시 시스템

Abstract

This paper proposes various ways to de-identify the original image region of the object so that it cannot be recognized, after verifying facial image region in real time by using YOLOv5 object recognition deep learning model in order to build intelligent surveillance system. The detection area is de-identified through an XOR operation with a noise file consisting of random pixel values, so that the face area required for decoding is not separately stored. The de-identification-completed image may be decrypted into the original image by performing the XOR operation again if necessary by the investigative personnel. The experiment was conducted with images of multiple people or fast-moving people. The results were well detected and de-identified in most situations, except in facial areas that were so small that they could not be visually identified, or in dark areas. The de-identification rate was measured to be about 25 fps, which can be used in real time.

Key Words : XOR operation, Noise file, Detecting face wearing mask, Face de-identification, Intelligent surveillance system

1. 서 론

국내 CCTV 설치 및 운영 현황[1]에 따르면 CCTV설치 대수는 계속 증가하고 있으며, 이와 더불어 블랙박스와 상업용 보안촬영카메라와 같은 실시간 촬영 장비들이 늘어남에 따라 사생활 및 초상권 침해의 문제를 일으키고 있다. CCTV로 상황을 확인함과 동시에 안면 유출 문제를 해결하기 위해서는 촬영 영상에 안면 비식별화 기술이 적용되어야 한다.

안면 객체의 범위는 현재 마스크 착용이 강제되는 국내외 상황을 고려하며, 마스크를 썼음에도 눈 영역을 통해 사람을 특정하는 경우가 생길 수 있으므로, 마스크를 착용한 얼굴을 포함시킨다.

이와 더불어 안면 비식별화 기술은 CCTV에 적용될 수 있도록 실시간 속도로 처리되어야 한다. 실시간 처리[2]라 함은 초당 사용되는 frame rate가 일정수준(24hz) 이상으로 유지되어 끊김이 없는 부드러운 동영상으로 인지될 수 있는 동작을 말한다.

또한 24시간 가동되는 CCTV의 특성으로 인해 매 순간 저장되는 영상의 용량을 조금이라도 아껴야 하는 필요성이 있다. 이를 위해 하나의 Noise파일을 만들어 기존 얼굴 영역과의 XOR연산을 진행하여 암호화를 하고, 작은 저장 공간을 사용하여 효율적으로 자동 비식별화 및 복호화를 한다. 단, 복호화 시 비식별화한 영상에 대해 연산을 적용하므로 비식별화한 영상은 lossless codec을 사용하여 원본 촬영 영상이 그대로 유지되어야 한다.

2. 관련 연구

기존 연구들은 대부분 단순 이미지 및 촬영된 저해상도 비디오에 위주로 암·복호화 연구[6]가 이루어졌고, 실시간으로 촬영되는 영상에 대해서 객체 영역을 비식별화하는 연구는 암호화 알고리즘을 사용한 비식별화 기법을 사용한 연구[7]에서만 찾아볼 수 있었다.

해당 연구[7]는 Advanced Encryption Standard (AES) 암호화 알고리즘이 낮은 연산 속도 때문에 편리하지 않다고 지적하며[8], multi-tasked cascaded convolutional neural network (MTCNN) 딥러닝 모델을 사용하여 안면 영역을 찾아내고, Reversible Chaotic Masking (ReCAM)이라는 새로운 암호화 기법을 제시하였다. 해당 모델은 실시간 검출 및 비식별화를 위해 엣지 컴퓨팅을 사용하였으며, 암호화 프로세스는 GTX1080ti와 벤치마크 결과가 비슷한[9] 라즈베리 파이4 환경에서 5.61fps의 낮은 처리속도를 보인다. 본 논문에서는 엣지 컴퓨팅과 필요 저장 용량의 추가 없이 빠른 속도로 비식별화를 하는 연구에 대해 논한다.

3. 모델 선정 및 안면 영역 검출

안면 영역을 비식별화 하기 위해서는 먼저 객체 검출(Object Detection) 과정이 필요하다. 현재 사용되고 있는 객체 검출 기법들은 one-stage detector와 two-stage detector로 나눌 수 있다. two-stage detector는 region-proposal과 classification과정이 동시에 진행된다. 대표적인 two-stage detector로는 R-CNN계열 모델이 존재한다. 반면 one-stage detector는 위의 두 단계를 통합하여 동시에 수행한다.

two-stage detector는 one-stage detector계열 모델에 비해 대체적으로 높은 정확도를 자랑하지만, 높은 계산 복잡도 때문에, 30fps를 넘는 운용이 불가하다는 단점이 있다[3]. CCTV에 적용할 객체 검출 기법은 CCTV의 특성을 고려하여 실시간으로 이루어져야 하므로, 현재까지의 다양한 모델들의 객체 검출 속도를 고려하면 one-stage detector 모델이 사용되어야 한다.

여러 객체 검출 모델들의 BOX AP(Average Precision) 순위는 COCO데이터 셋으로 행한 벤치마크 테스트(2019.06)[4] 에서 다음과 같은 결과를 확인할 수 있다. 아래 그림 1에서는 EfficientDet-D7모델이 높은 AP 수치를 나타낸다.

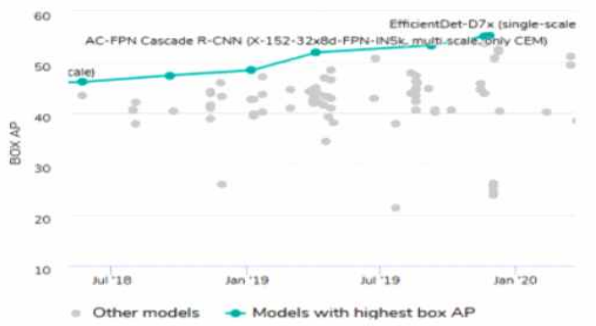

Fig 1. COCO test-dec Object Detection Ranking[4]

아래 표 1은 EfficientDet모델과 다른 객체인식 모델들 간의 AP수치의 차이를 보여주고 있다. EfficientDet모델은 one-stage 모델로 상대적으로 적은 연산량(FLOPS)으로 높은 AP를 보여준다. 하지만 GPU RTX2060의 환경에서 평균 FPS가 1.5로 매우 느린 속도를 보였다는 연구 결과가 있다[5]. 실시간 비식별화 기능을 수행하기 위해서는, AP는 EfficientDet과 비슷하고 높은 FPS의 성능을 가진 모델이 필요하다.

Model	BOX AP	AP50	AP75
EfficientDet-D7x	55.1	74.3	59.9
EfficientDet-D4	49.7	68.4	53.9
Mask R-CNN	46.1	60.5	44.1
EfficientDet-D0	34.6	53.0	37.1
Faster R-CNN	34.7	63.6	46.4
YOLOv3	43.9	64.1	49.2
SSD	28.8	48.5	30.3

Table 1. Comparison of Oject Detection Models[4]

YOLOv5모델은 대표적인 one-stage detector로서, 아래 그림5에서 확인할 수 있듯이 AP값은 EfficientDet과 비슷하지만 약 5배가량 빠른 성능을 보여준다. 본 논문에서는 YOLOv5l 모델이 사용되었다.

해당 모델을 사용하여 안면영역을 검출한다. 검출 결과로 나온 영역의 왼쪽 위 좌표와 오른쪽 아래 좌표가 해당 영역을 비식별화 하는데 사용된다.

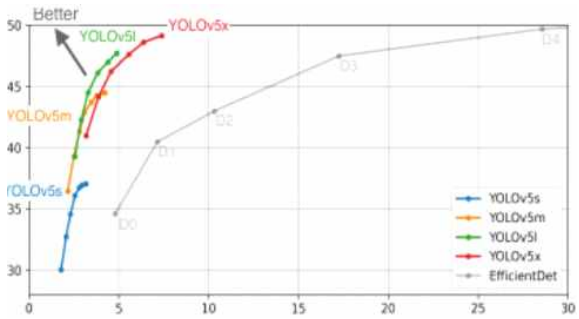

Fig 2. Comparison Between YOLOv5 Model and EfficientDet

4. 제안한 비식별화 및 복호화 방안

딥러닝 모델을 사용해 검출된 안면 영역들을 실시간으로 비식별화 할 수 있는 방법들은 다양하게 존재한다.

4.1 암호화 알고리즘

암호화 알고리즘은 각 알고리즘에 따른 성능 차이가 약간씩 존재하지만, 대체적으로 해독이 불가하여 보안성이 뛰어나다. 얼굴 영역을 암호화 후 얼굴 영역을 따로 저장하고, 복호화를 하기 위해 저장한 얼굴 영역을 이용하는 암호화 방식은, 그만큼 저장용량이 추가로 필요하기 때문에, 해당 얼굴 영역만큼의 저장 용량이 필요하다는 단점이 있다. 반면 기존 암호화 알고리즘을 사용한 연구들[7,8]은 저장 용량에 대한 문제는 없지만, 낮은 연산속도 및 처리 속도를 보여주었다.

4.2 노이즈 파일과 기존 얼굴 영역의 각 픽셀 값들을 비율을 정해 혼합하여 저장

무작위 값으로 생성된 픽셀들로 이루어진 노이즈 파일과 변환할 기존 얼굴 영역의 픽셀 값들을, 비율을 정하여 혼합하여 저장하는 방식이다. 이 방식은 랜덤한 픽셀 값으로 이루어진 노이즈 파일을 따로 생성하여, 이 파일 하나만으로 비식별화와 복호화를 진행하기 때문에, 3.1의 방식과는 달리 간단한 암호화 방식으로서 빠른 연산속도와 얼굴영역을 저장하는 용량이 따로 필요하지 않다는 장점이 있다.

얼굴 영역과 노이즈 파일을 2:8 비율로 혼합하였을 때, 기존의 얼굴 윤곽이 미세하게 보이기 때문에, 아래의 수식 1과 같이 얼굴영역과 노이즈 영역 픽셀 혼합 비율을 1:9로 노이즈 파일의 비율을 높여주었다. 두 이미지의 픽셀들을 혼합하는 코드는 cv2 라이브러리의 addWeighted함수를 사용하였다.

$$encrypted = cv2.addWeighted(face,\ 0.1,\ noise,\ 0.9,\ 0.0_{(scalar\ added\ to\ each\ sum)})$$

Fig 3. mixing two image region

이 방식의 문제점은 암호화와 복호화 시점 모두에서 나타난다. 암호화 시 안면 영역의 픽셀과 노이즈 파일의 픽셀에 각각 할당된 비율인 소수점 자리 숫자를 곱하고 더하는 과정에서, 남겨진 소수점 자리 숫자가 버려진다. 복호화 시 다시 배수를 곱하는 과정에서 원본 영상과의 픽셀값 차이가 커지게 된다.

4.3 노이즈 파일과 기존 얼굴 영역의 각 픽셀값끼리 XOR 연산을 통해 암호화 및 복호화

4.2의 방법을 수행했을 때 원본 영상과의 픽셀 값 차이가 피치 못하게 발생하기 때문에, 해당 방법은 사용할 수 없다고 결론을 내렸다. XOR연산(1)의 가장 큰 특징은, 같은 연산을 2회 반복하게 되면 원래의 값으로 돌아온다는 것이다.

$$(P \oplus K) \oplus K = P \quad \cdots\cdots (1)$$

위의 특징을 사용해, 노이즈 파일의 픽셀값과 기존 얼굴 영역의 픽셀값을 XOR연산 하여 비식별화를 진행하고, 복호화가 필요하다면 같은 XOR연산을 반복하면 된다. 이 방법의 장점은 연산 후 클램핑이 일어나는 픽셀값들이 존재하지 않고, 암호화와 복호화의 연산과정이 단순하고 빠른 점이다. 또한 4.2방법에서 얼굴 영역을 따로 저장하지 않아도 되는 장점이 그대로 유지된다.

프로세스 전체 과정은 아래 사진 5와 같다. 사진 5의 XOR operation은 3차원 배열인 noise와 face영역에 대해 연산을 진행하는 것이며, 해당 연산 과정은 사진 4와 같다. 수식에서 encrypted는 암호화된 픽셀값이며 해당 연산

은 face영역 전체의 픽셀에 대해 적용된다.

```
for i=0 to len(face) do
    for j=0 to len(face[0]) do
        for k=0 to range(3) do
            encrypted=face[i][j][k]^noise[i][j][k]
```

Fig 4. XOR operation between face region and noise file

결과적으로 4.1 방법과의 용량 차이는 이미지 프레임 내에서 검출되는 얼굴 영역의 크기의 합 * 프레임 수와 같다.

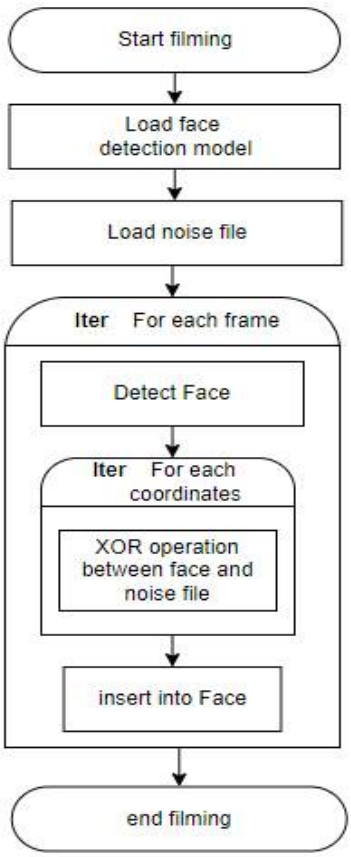

Fig 5. Flowchart of proposed encryption process

4. 시뮬레이션 및 결과

4,1 XOR연산 사용 결과

최종적으로 제안된 방법은 랜덤 픽셀 값들로 이루어진 노이즈 파일을 생성한 뒤, 해당 파일의 각 픽셀 값과 촬영되는 영상의 프레임의 각 픽셀 값과 XOR 연산을 통해 비식별화 및 복호화 하는 방법이다. 이 방법을 통해, 비식별화 및 복호화 과정에서 불필요한 오차 값이 생기지 않고, 복호화 영상은 깨끗하게 원본 픽셀값이 도출된 것을 아래 그림4에서 볼 수 있다. 또한 프레임마다 검출한 얼굴 객체들의 사진을 따로 저장할 필요 없이, 노이즈 파일 하나로 저장 공간을 줄일 수 있게 되었다.

비식별화된 얼굴영역과 복호화된 얼굴영역은 그림 6의 (a)와 (b)로 확인할 수 있다. 안면인식 및 비식별화 알고리즘은 640(w)*480(h)*3(rgb) 영상에 대해 평균 0.025s의 속도가 소요되었다.

Fig. 6. Result of de-identified face region and restored region: (a) de-identified state (b) restored state

사진 7과 8은 1280*720 크기의 동영상에 대해서 실험한 결과이다. 사진 7의 다수의 사람들에 대해서 실험을 했을 때는, 사람의 숫자와 상관없이 측정이 잘 이루어졌지만, 검출되는 안면영역이 약 16*20 픽셀보다 작아지면 측정이 되지 않는 결과를 보여주었다. 하지만 이는 육안으로 확인이 불가할 정도의 크기이기 때문에 한계점으로 판단하지 않았다.

Fig 7. An experiment that measures the minimum pixel area and performance when given a large number of people

속도에 따른 성능을 측정을 하였을 때는, 전동킥보드를 타는 사람 및 약 시속 25km정도로 헬멧을 쓴 채 빠르게 움직이는 오토바이를 타는 사람의 옆모습에 대해서도 측정이 잘 이루어졌다.

Fig 8. A speed measurement experiment conducted on a person riding a motorcycle

4.2 한계점

안면 인식 모델의 한계점을 알아보기 위해 빛의 양(lux)에 따라 얼굴 인식도를 측정하였다. 측정 결과로 lux값이 60이하로 내려가면서부터 검출이 되지 않는 프레임이 생겼다.

다음으로 얼굴 각도에 따라 얼굴 인식도를 측정하였다. 이때 실험은 220 lux 환경에서 진행하였다. 해당 환경에서 안면 영역을 검출하지 못한 경우는 사진 9처럼 고개를 숙인 후 눈이 보이지 않는 경우와, 사진 10처럼 약 170도로 뒤 돌은 경우가 있다. 이외의 경우에는 모두 검출이 되어 비식별화가 완료되었다.

Fig 9. Face with bent down pose

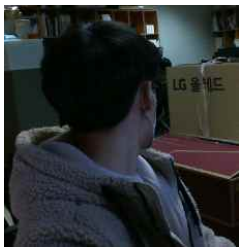

Fig 10. Face turned around more than 170 degrees

딥러닝 모델의 한계점 이외에, 복호화에는 XOR연산에 필요한 픽셀값들을 정확하게 유지하기 위해서 lossless compression을 위한 영상 코덱이 필요했는데, 해당 코덱을 사용하게 되면 비식별화 후의 영상 저장 용량이 급격하게 느는 단점을 보였다. 하지만 복호화 후의 영상 크기는 lossy compression codec들을 사용할 수 있기 때문에 비식별화 후의 영상은 저장 대신 화면 송출 정도로만 쓰일 수 있을 것으로 보인다.

5. 결론 및 향후 연구

본 논문에서는 CCTV 화면상에 나오는 모든 얼굴들을 효과적으로 자동 비식별화 및 복호화 하는 방법을 제안하였다. 제안된 방법은 랜덤 픽셀 값들로 이루어진 노이즈 파일을 생성한 뒤, 해당 파일의 각 픽셀 값과 촬영되는 영상의 프레임의 각 픽셀 값과 XOR 연산을 통해 비식별화 및 복호화 하는 방법이다. 이를 통해 프레임마다 검출한 얼굴 객체들의 사진을 따로 저장할 필요 없이, 노이즈 파일 하나로 저장 공간을 줄이며, 빠른 속도로 암호화 프로세스를 처리할 수 있게 되었다.

다만 영상 저장 코덱에 따라 픽셀값이 달라지는 점을 고려하여, 복호화를 하기 위해 비식별화 단계에서 무손실 코덱을 사용하여 영상 저장을 해야 하기 때문에, 영상 저장 용량이 많이 필요한 한계점을 보였다. 또한 노이즈 파일 한 장으로 비식별화와 복호화 둘다 이루어지기 때문에 보안에 취약할 수 있다는 단점이 있다.

또한 얼굴 검출 시 빛이 희미하게 있는 광원이 부족한 상황과, 얼굴 각도가 일정 각도 이상 틀어졌을 시, 어두운 배경에서 피부색이 어두운(흑인) 안면 식별 시 제대로 작동하지 않는 경우도 발생하였다. 광원이 부족한 사진을 모아놓은 데이터 셋을 구하지 못해 생겨난 한계점으로 생각된다.

향후 연구에서는 무손실 코덱을 사용하지 않거나, 효율적으로 영상을 압축하는 코덱에 대한 연구가 필요할 것으로 보인다. 또한 노이즈 파일을 사용하는 것으로 생기는 보안 취약점은 서버와 클라이언트(촬영자) 관계로 공개키와 개인키를 추가하여 노이즈파일을 암호화하는 등의 방식으로 보안성을 향상시킬 수 있을 것으로 보인다.

References

[1] 통계청, “공공기관 CCTV 설치 및 운영,” Available: https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=2855, (2012. December 20)

[2] 이석호, “디지털 동영상의 프레임레이트와 모션블러가 수용자의 시지각에 미치는 영향에 관한 연구,” 한국과학예술포럼, vol. 12, 2013, 175 - 184.

[3] Chun, Dayoung, Jiwoong Choi, Hyun Kim, and Hyuk-Jae Lee, “A study for selecting the best one-stage detector for autonomous driving,” In 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), pp. 1-3, 2019.

[4] Paperswithcode, “Object Detection on COCO test-dev,” [Internet]. Available: https://paperswithcode.com/sota/object-detection-on-coco. (2012. December 20)

[5] 이지원, “최적화된 차량 탑승인원 감지시스템 개발을 위한 딥러닝 모델 분석,” 한국정보통신학회논문지, vol. 25, 2021, 146 - 151.

[6] Ribaric, Slobodan, and Nikola Pavesic, “An overview of face de-identification in still images and videos,” In 2015 11th IEEE International Conference and Workshops on Automatic

Face and Gesture Recognition (FG), vol. 4, pp. 1-6, 2015.

[7] Fitwi, Alem, Yu Chen, Sencun Zhu, Erik Blasch, and Genshe Chen, "Privacy-preserving surveillance as an edge service based on lightweight video protection schemes using face de-identification and window masking," Electronics 10, no. 3, 2021, pp.236.

[8] Liu, Lingfeng, and Suoxia Miao, "A new image encryption algorithm based on logistic chaotic map with varying parameter," SpringerPlus, vol 5, no. 1, 2016, pp.1-12.

[9] Roel P. "Raspberry Pi 4: GPU speed," [Internet]. Available: https://www.cloudsavvyit.com/8239/raspberry-pi-4-good-enough-for-gaming/, (2012. December 20)