영상 데이터 개인정보 비식별화 기법 연구

송인준, 김차종*

한밭대학교 정보통신전문대학원, *한밭대학교 컴퓨터공학과

ijsong@watosys.net, cjkim@hanbat.ac.kr

Video Data Personal Information De-identification Techniques

Injun Song, Chajong Kim*

Department of Computer Engineering, Graduate School of Information & Communications, Hanbat National University

Department of Computer Engineering, Hanbat National University*

요 약

본 논문은 영상 데이터에 대해 개인정보를 보호하기 위해 기검출된 개인정보 영역에 대한 보안성과 재사용성이 높은 개인정보 영역 비식별화 방법을 제안한다. 비식별화 방법으로는 셔플링 기반의 마스킹 방법을 사용하였고 마스킹 정보에 대한 이중키 기반의 암호화를 통해 타인에 의해 복원할 수 없도록 하였으며, 원본 영상의 재사용을 위해 보안키를 통한 원본 영상 복원이 가능하도록 하였다. 본 논문의 연구 결과는 영상 데이터에 포함된 개인정보 영역에 대한 비식별화를 통해 개인정보 유출이 없는 데이터의 산업적 활용과 영상을 활용한 산업 분야에서 개인정보 보호에 드는 비용을 줄이는데 기여할 것으로 기대한다.

Ⅰ. 서 론

최근 인공지능의 한 분야인 딥러닝 기술의 발달에 따라 자율주행과 로봇, 스마트폰 응용 등의 분야에 많이 활용되고 있다. 이러한 인공지능 기술이 우리 생활에 더욱 밀접하게 확장되면서, 인공지능이 자동으로 데이터를 수집하고 학습하는 경우가 증가하고 있다[1]. 특히, 자율주행차나 로봇 분야에서는 실제 환경의 인공지능 학습데이터를 위해 많은 양의 영상 데이터를 수집하고 있으며, 개인정보(얼굴, 차량 번호판 등)가 포함되어 있어서 법적으로 많은 논란이 되고 있다[2]. 또한, 공공장소와 일반 생활공간에서 수많은 영상이 수집되고 이용되고 있으며, 무분별한 영상 데이터의 수집으로 인해 개인정보가 유출되는 피해가 속출하고 있다. 일 예로, 국내외에서는 안전을 위해 많은 수의 CCTV를 설치하고 활용하고 있으며, 지능형 관제 시스템을 도입함으로써 기존에 수집된 데이터를 학습용 데이터로 가공하면서 개인정보 유출의 위험이 발생하게 된다[3][4]. 개인정보가 포함된 영상 데이터는 법적 요구에 따라 개인정보 영역에 대해 마스킹 처리 등과 같은 비식별화를 거쳐 저장하고 관리, 이용, 전송되어야 한다. 자율주행차, CCTV 등에서 촬영한 영상 데이터의 비식별화는 법적으로 요구되는 매우 중요한 과정이다[5]. 본 논문에서는 영상 데이터에 대해 개인정보를 보호하기 위해 인공지능 기반으로 미리 검출된 개인정보 영역에 대한 보안성과 재사용성이 높은 개인정보 영역 비식별화 방법을 제안한다.

Ⅱ. 영상 데이터 개인정보 영역 비식별화 처리 방법

1. 영상 데이터 개인정보 영역 마스킹 원리

개인정보 영역은 사람과 컴퓨터가 알아볼 수 없는 형태로 데이터 마스킹 처리를 하여야 한다. 또한 마스킹 처리된 영상은 제3자가 쉽게 변형할 수 없어야 한다. 이를 위해 본 논문에서는 마스크 테이블을 랜덤하게 생성하고 키값을 통해 마스크 테이블을 추출하는 방법과 테이블을 통해 픽셀의 위치를 교환하는 2단계의 하이브리드 방법을 통해 마스킹 처리하는 방법을 제안하고자 한다. 마스킹 키값과 마스킹 처리된 개인정보 영역에 대한 좌표값은 대칭키 암호화 알고리즘으로 암호화된다. 스크램블링(Scrambling) 기법을 사용한 기존 연구에서는 무작위 난수 발생을 통해 영상의 암호화 마스킹 처리를 하고 있으나, 실용성 측면과 해킹 위험성 측면에서는 사용자가 마스킹 키와 셔플링 테이블을 직접 관리하는 것이 보안과 관리 측면에서 위험성을 낮추는 방법이다[6][7][8].

2. 영상 데이터 개인정보 영역 마스킹 처리 절차

영상 데이터 비식별화 단계는 두 단계로 구분할 수 있다. 첫 번째는 영상 내에서 개인을 식별할 수 있는 영역을 탐지하는 단계이고, 두 번째는 탐지한 개인 식별 영역을 변형하는 단계이다. 국내외에서는 영상에서 개인을 식별할 수 있는 영역에 여러 필터를 적용하여 특정한 개인을 식별하지 못하게 하는 기술과 영상을 암호화하여 허가된 대상에게만 공개하는 방법, 영상 데이터에 적합하게 K-익명성 모델을 확장하여 수집된 얼굴 영상 내에서 K개의 얼굴을 합성하여 적용하는 방식으로 개인을 식별할 확률이 1/K를 넘지 않게 하는 기술이다.

마지막 방법은 영상에서 특정한 부분을 제거하고 공백이나 손상된 부분을 채우는 방법이다[9].

Ⅲ.실험

1. 실험 준비

본 논문에서는 인마스킹 처리하기 위해 셔플링(shuffling) 정보 데이터를 찾아올 키를 지정하고, 개인정보 영역을 식별할 수 있는 바운딩 박스 정보를 저장할 메타데이터에 대한 암호화 키를 지정한다. 그다음 영상 데이터에서 개인정보에 해당하는 사람의 얼굴과 차량 번호판을 추출하고, 위에서 지정한

* Corresponding Authors: Chajong Kim(cjkim@hanbat.ac.kr)

셔플링 키값을 기반으로 마스킹 처리를 수행할 테이블값을 추출한 다음, 개인정보 영역에 대한 마스킹을 수행한다.

개인정보 영역에 대한 마스킹 과정은 무결성 확보를 위해 2번의 데이터 셔플링이 이루어진다. 그림 1은 개인정보 비식별화 처리 과정이다.

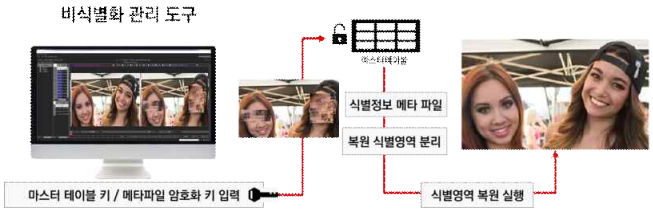

그림 1 마스터 키값 기반의 개인정보 비식별화 상세 흐름도

이때 비식별화는 그림 2와 같이 마스터 테이블 킷값을 통한 보안 기술이 적용된 셔플링 기법을 적용하였으며,보안 기술은 1,024개의 8X8 마스크 중에서 랜덤하게 찾아오는 방법이 담긴 유일한 키값을 사용자가 지정하고, 개인정보 검출 영역에 대한 셔플링을 수행할 개인정보 영역에 대한 좌표값을 암호화할 키값을 지정하여 암호화된 비식별화 방법을 의미한다. 좀더 자세히 설명하면 영상 데이터에서 검출한 비식별화 대상 영역에 대한 정보는 식별정보로서 메타데이터의 형태로 저장된다. 영상을 복원할 때는 저장된 메타데이터를 기반으로 키값을 적용하여 셔플링 테이블을 찾은 다음 복원 과정을 수행한다.

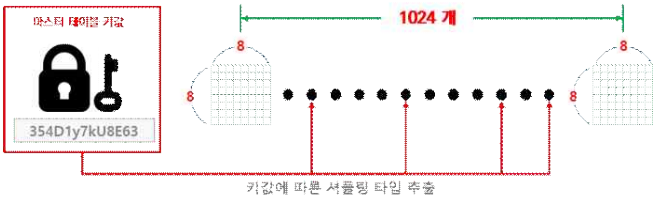

그림 2 테이블 키값 기반의 셔플링 규칙 추출과정

1,024개의 8X8 마스크는 랜덤하게 생성된 셔플링 규칙을 담고 있으며, 마스터 테이블 키값을 기반으로 1,024개의 후보 중에서 추출한다. 예를 들어 하나의 영상에서 개인정보 영역이 3개 검출되었다면, 키값을 기반으로 마스크는 3개가 추출되고 각 영역은 서로 규칙으로 셔플링된다.

셔플링 테이블을 구하는 키 값은 랜덤 함수를 통해 추출하며, 수식 (1)와 같이 C++의 랜덤함수를 활용할 수 있으며,

$$Key\ Value = rand(Seed) \cdots\cdots\cdots\cdots (1)$$

1,024개로 구성된 셔플링 규칙 세트는 이론적으로 팩토리얼 64 만큼의 후보 중에서 선정됨으로 유일한 세트는 무한대에 가깝다고 할 수 있다. 본 논문의 마스킹 방법을 적용하면 사용자의 선택에 따라 1,024개의 세트를 선택할 수 있고, 마스크 처리할 테이블은 키값을 기반으로 무작위로 선정하게 되므로 강력한 보안성능을 가진다. 그림 3은 마스킹 적용 방법에 대한 예시이다.

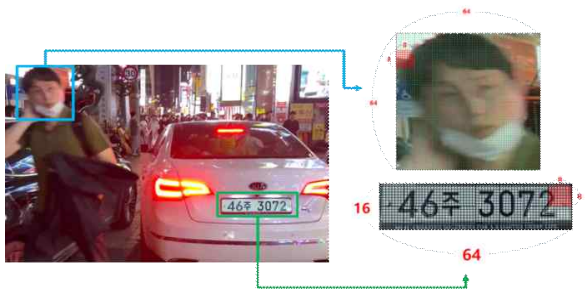

그림 3 8X8 마스크를 이용한 셔플링 방법

2. 실험

그림4와 같이 개인정보 영역에 대한 본 논문에 제시한 방법으로 마스킹 처리 후 객체 검출률을 측정해보며 객체 검출이 안되는 것을 확인할수 있다. 또한 그림 5에서와 같이 보완키값에 의해 복원된 이미지의 객체 검출률을 살펴보면 원본 이미지의 객체검출률과 동일하게 측정되어 원본으로 복원이 잘 되었다는 것을 확인할 수 있다.

그림 4 원본 이미지와 마스킹 이미지의 객체 검출률 비교

그림 5 원본 이미지와 복원된 이미지 객체 검출률 비교

Ⅲ. 결론

영상 복원을 쉽게 하면서도 보안을 유지할 수 있는 대칭키 기반의 하이브리드 마스킹 기술 기반의 비식별화 처리 방법을 통해 차량 번호판 83%, 사람 얼굴 91%의 검출률을 가지는 원본 이미지의 개인정보 영역이 하이브리드 마스킹 후 검출률을 측정했을 때 두 영역 모두 0%로 검출되지 않았고, 대칭키를 이용하여 복원된 이미지의 개인정보 영역에서 원본 이미지와 동일한 차량 번호판 83%, 사람 얼굴 91%의 검출률을 얻을 수 있었다. 이를 통해 영상 데이터의 보안성과 활용성을 확인할 수 있었다.

참 고 문 헌

[1] 김윤정, 윤혜선, “인공지능 기술의 활용과 발전을 위한 제도 및 정책 이슈”, 한국과학기술기획평가원, ISSUE PAPER 2016-07, pp.3-30, 2016.

[2] 민경욱, 최정단, “자율주행 인공지능 기술 개발을 위한 데이터 수집 및 학습 플랫폼”, 2020년도 한국자동차공학회 춘계학술대회, 2020.

[3] 장인호, “첨단과학기술사회에서 이동형영상정보처리기기 이용 확산에 따른 개인영상정보 보호방안에 관한 연구”, 성균관대학교 법학연구원, 제28권, 제2호, pp.31-78, 2016.

[4] 홍준혁, 이병엽, “인공지능기반 보안관제 구축 및 대응 방안”, 한국콘텐츠학회, 제21권, 제1호, 2021.

[5] 원병철, “어린이집 CCTV 영상 반출하려면 모자이크 필수... 비용 문제는 해결 못해”, 보안뉴스, 2021, 04.23., https://www.boannews.com/media/view.asp?idx=96831

[6] 이동혁, 박남제, “지능형 영상 감시 환경에서의 개인정보보호를 위한 COP-변환 기반 메타데이터 보안 기법 연구”, Journal of The Korea Institute of Information Security & Cryptology VOL.28, NO.2, Apr. 2018

[7] F. Dufaux, T. Ebrahimi, “Scrambling for Privacy Protection in Video Surveillance Systems”, IEEE Trans. on Circuits and Systems for Video Technology, Vol. 18, No.8, pp.1168-1174, 2008.

[8] F. Dufaux, T. Ebrahimi, “A Framework for the Validation of Privacy Protection Solutions in Video Surveillance”, IEEE International Conference on Multimedia and Expo (ICME), pp. 66-71, 2010.

[9] 문용식, “영상 데이터 익명화 기술 및 평가방안”, 한국정보화진흥원, 제11호, 2019.