

인공지능 기술을 활용한 사생활 보호 기능 탑재 CCTV 카메라의 설계 및 구현

김성규, 문남미
호서벤처대학원 융합공학과
20225476@vision.hoseo.edu, nmm@hoseo.edu

Designing and Implementing a Privacy Protective CCTV Camera Using Artificial Intelligence Technology

Sungkyu Kim, Nammee Moon
Dept.of Convergence Engineering, The Graduate School of Venture

요 약

CCTV 시스템은 다양한 목적으로 매우 많은 곳에서 유용하게 활용되고 있으나, 해킹 또는 기타의 이유로 사생활 침해에 대한 우려가 커지고 있다. 본 논문에서는 영상의 입수 시점에서부터 사생활 침해가 발생할 수 있는 영상 정보를 제거하고, 대신 인공지능 기술을 활용하여 이를 대체할 수 있는 인물의 스펙트럼 정보를 추가로 제공함으로써 사생활 정보를 침해하지 않으면서도 상황을 충분히 인식할 수 있는 CCTV 용 카메라 시스템을 여러 인공지능 알고리즘을 비교하여 설계하였다. 또한 카메라 및 인공지능 기술을 쉽게 사용할 수 있으면서도 주변에서 쉽게 구할 수 있는 안드로이드 스마트폰을 이용하여 설계를 실제 구현한 후 영상의 각 해상도별 초당 처리 성능 및 스펙트럼 인식률 등의 성능을 평가하여 실제로 제품구현 및 제품화 시 중점을 두어야 할 주요 개선점 및 향후 연구과제를 도출하였다.

1. 서론

CCTV(Closed Circuit Television, 폐쇄회로텔레비전)는 외부의 네트워크와 연결되지 않은 폐쇄망에서 사용되는 영상획득, 녹화, 시청이 가능한 장치 또는 시스템을 말하며, 다양한 분야에서 안전 확보 및 사후 처리를 위하여 널리 사용되고 있다[1]. 공익적인 목적으로 범죄 발생 시 결정적인 증거로 활용되고 있으며 [2], 최근에는 인공지능 기술을 활용하여 자동으로 이상행동을 파악 범죄 행위를 선제적으로 경고하거나[3] 또는 문제상황을 자동으로 인지하여 경보를 발생시키는 스마트시스템이 활용되고 있으며, 사람들의 행동 패턴을 분류, 행동을 예측하여 상품을 제안하는 등의 마케팅 영역, 외출 시 거동이 불편한 사람 또는 반려동물이 있는 집안을 관찰하는 용도로도 널리 활용되고 있다.

그러나, 중국의 경우처럼 자국민을 감시하는 데 사용되기도 하며, 개인의 의사에 무관하게 촬영이 되는 것을 사생활 침해라고 생각하는 사람들이 늘어나고 있다. 또한 개인정보보호법상으로도 심각한 사생활 침해 우려가 있는 목욕실, 화장실, 발한실(發汗室), 탈의실 등의 장소에는 설치가 금지되어 있다[4]. 최근에는 다양한 시스템, 네트워크와 연결되면서 전 세계적으로 해킹 피해가 발생하고 있다.

본 연구에서는 인공지능 기술을 활용하여 영상을 취득단계에서 분석 및 변조 처리하여 사생활정보를 제거하고, 누락, 변조된 대상 인물의 행동 정보를 스펙트럼 형태로 제공함으로써 원본 영상을 제공하지 않으면서도 상황 인지를 명확히 할 수 있는 시스템을 설계 및 구현하고, 그 결과를 분석하여 주요 개선필요점 및 향후 연구과제를 분석하였다.

2. 카메라 시스템의 설계 및 기술적 배경

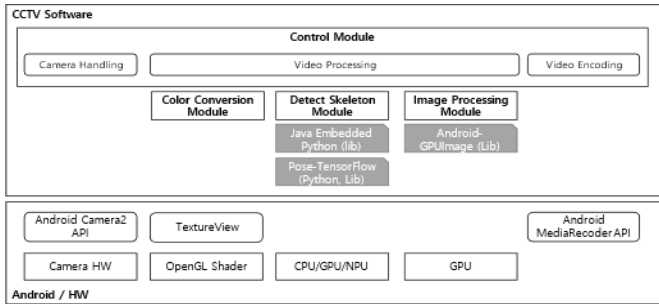
일반적인 CCTV 카메라 시스템의 구조는 (그림 1)과 같이 영상취득 용 CCD 또는 CMOS 센서부와 센서에서 영상을 압축하는 Codec 부, 압축된 영상을 외부로 전달하는 외부 인터페이스부, 그리고 모든 모듈들을 컨트롤하는 제어 소프트웨어부로 구성된 구조이다.



(그림 1) 일반적인 디지털 CCTV 카메라의 구조

본 논문에서는 (그림 2)와 같이 제어 소프트웨어부를 개선하여 영상 취득부와 압축부 중간에 영상에서 사생활정보를 변조, 제거하고 이를 보완하는 스펙트럼

톤 정보를 추출하고, 스켈레톤 정보를 추가하는 영상 처리 모듈을 추가하는 방식으로 설계를 변경하였다.



(그림 2) 사생활보호 기능을 탑재한 카메라 시스템의 구조

영상에서 사생활 정보를 제거하기 위해 영상을 변조하는 방법은 오래전부터 개발되어 널리 사용되어 온 블러(Blur)[5] 또는 모자이크(Mosaic)[6] 필터를 적용하였다.

사람의 자세를 탐지하여 스켈레톤 형태의 정보를 추출하는 방법은 크게, 영상 속에서 사람을 인식하고 해당 영역에서 사람의 자세를 인식하는 Top-down 접근방법과 영상 속에서 사람의 개별 부위(Heat point)를 우선 인식하고 이들 집합에서 사람으로 조합 가능한 세트를 찾아내는 Bottom-up 접근방식 두 가지로 나누어진다. Top-down 방식의 경우 두 번의 추론 과정을 거쳐야 하지만, Bottom-up 방식의 경우 추론이 한 번만 필요하기에 상대적으로 연산의 양이 적고, 사람의 일부가 가려진 경우 더 나은 인식성을 보이며, 등장인물이 다수의 등장인물을 모두 인식이 가능하면서도 등장인물의 수에 따른 성능의 차이가 작아 최근의 연구 방향은 Bottom-up 방식을 위주로 연구가 진행되고 있다. Top-down 방식의 대표적인 알고리즘으로는 AlphaPose[7] 알고리즘이 있으며 Bottom-up 방식의 대표적인 알고리즘으로는 DeepCut[8] 알고리즘과 이를 개선한 DeeperCut[9], ArtTrack[10] 알고리즘 등이 있다.

3. 사생활보호 기능을 탑재한 카메라 시스템의 구현

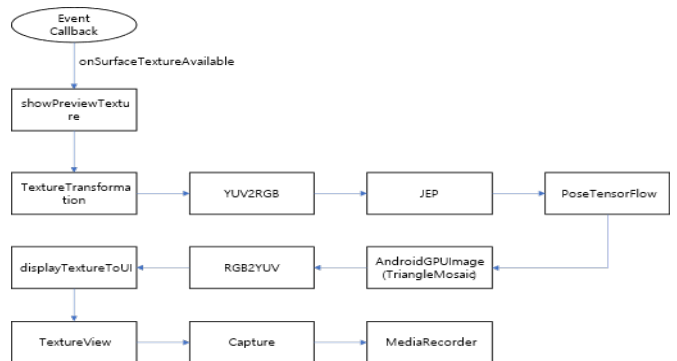
본 연구에서는 쉽게 인공지능 기술을 활용할 수 있고, 카메라를 포함하며 쉽게 구할 수 있는 안드로이드 휴대폰을 활용하여 연구를 진행하였다.

영상의 취득은 안드로이드의 Camera2 API 를 활용하였고, 다시 영상을 압축하는 코덱은 MediaRecorder API 를 활용하여 최종적으로 MP4 스트림 형태의 영상을 생성하도록 구현하였다.

영상 처리 모듈은 카메라 모듈에서 입수된 YUV 컬러 영상을 인공지능 및 영상처리 모듈에서 사용이 가능한 RGB 컬러 영상으로 변환한 후 스켈레톤 탐지 모듈로 전달하고 이후 사생활정보를 제거하는 모듈로 전달한다. 최종적으로 다시 YUV 컬러로 변환하여 코덱으로 전달한다.

영상의 컬러 체계인 YUV 형식과 RGB 의 변환을 위해서 많은 연산이 필요했으며, 성능이 대폭 저하되

는 현상이 발생하였다. 이의 개선을 위하여 안드로이드에서 사용가능한 방법 중 OpenGL shader 를 활용하여 개선하였다. <표 1>은 YUV/RGB 컬러 변환 방법 별 성능을 비교한 결과이다.



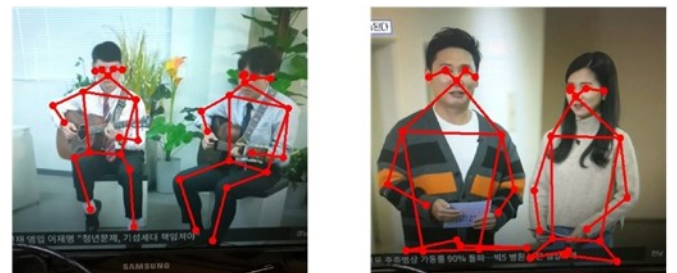
(그림 3) 영상 처리 프로세스

<표 1> YUV/RGB 컬러 변환 방법 별 성능

가속 방법	성능 (FPS)
CPU	1.2
Android RenderScript	7.1
OpenGL Shader	15.6

스켈레톤 탐지 모듈의 경우 DeepCut 알고리즘의 구현 프로젝트인 Pose-TensorFlow 프로젝트를 이용하였다. 그러나, 해당 라이브러리는 Python 으로 구현된 라이브러리 이기에 안드로이드에서 기본적으로 사용하는 언어인 Kotlin 에서 직접 사용할 수가 없어 Java Embedded Python 프로젝트를 활용하여 Kotlin 에서 호출할 수 있도록 하였다.

추론을 위한 자세 모델은 Google 이 개발하여 공개하였으며 지속적으로 업데이트 중인 Multi person 용 데이터셋 MoveNet MultiPose 모델을 활용하였으나, 안드로이드에서 사용하기 위해서는 TFLite Converter 도구를 이용하여 TensorFlow Lite 용으로 변환하여야 하였다.



(그림 4) 스켈레톤 탐지 결과

사생활보호를 위하여 영상을 변조하는 방법은 안드로이드 상에서 OpenGL shader 를 활용하여 빠르게 처리할 수 있는 Android-GPUImage 라이브러리를 활용하여 구현하였으며, 다양한 종류의 필터를 적용해봤으나 OpenGL 가속 기능이 없는 실제 구현 하드웨어 성능을 고려하여 간단하면서도 빠르고, 훌륭한 보호 성능을 제공하는 모자이크(Mosaic) 필터를 사용하였다.

4. 실험 및 평가

결과물의 테스트 장소는 형광 등으로 조명이 이루어지는 일반적인 사무실을 이용하였으며, 평가 기준은 3 가지 해상도 별 초당 처리된 프레임 수(FPS)와, 처리한 전체 프레임 수에 대한 스켈레톤 탐지 성공 비율인 자세 인식률 두 가지 값으로 측정하였다.

<표 2> 성능 측정 결과

해상도	평균 FPS	정확도 (%)
1920x1080	3.4	87.2
640x480	7.6	74.2
320x200	11.2	31.7

<표 2>는 그 결과를 정리한 표로 해상도가 높아질수록 탐지성공확률은 높아지나 처리해야할 데이터의 양이 증가하여 초당 처리 성능은 반대로 낮아지는 것을 확인하였다. 현재로서는 실제 활용이 가능한 최소한의 수준인 만화영화(12FPS)정도의 성능에도 모든 경우 미치지 못하는 것으로 확인되었다.

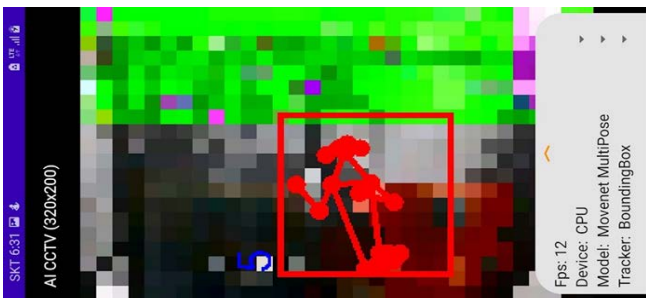
정성적으로는 320x200 의 해상도에서는 변조 이후 거의 상황인식이 불가능하였으며 640x480 의 해상도에서는 어느정도 상황의 인식은 가능했으나 스켈레톤 탐지율의 저하가 큰 편이라 1920x1080 이상 해상도의 영상을 활용하는 것이 적절하다고 판단되었다.



(그림 5) 1920x1080 해상도의 예



(그림 6) 640x480 해상도의 예



(그림 7) 320x200 해상도의 예

5. 결론 및 향후 연구과제

1920x1080 이상 해상도의 경우 사생활 보호를 제공하면서도 상황인식이 가능한 화면을 보여줄 수 있다는 것을 확인하였으나 큰 처리 성능의 개선이 필요하다는 것 또한 확인하였다.

가장 큰 연산량이 필요한 YUV/RGB 컬러 변환 과정을 OpenGL shader 를 활용 개선하였으나, 근본적으로 이 과정 자체를 생략할 수 있는 YUV 컬러 기반 인공지능 모델의 개발이 필요하고, 불필요한 라이브러리들을 제거하면 실제 제품화 가능한 성능을 확보할 수 있을 것으로 판단된다.

참고문헌

- [1] Hyunbin Cho, Hodae Cho, "Problem and Improvement Methods of CCTV Practical Use For Social Security", Korean Journal of Public Safety and Criminal Justice, vol. 19(1), pp259-283, 2010.
- [2] Kyungdon Kim, "On Collection and Management Process for the Admissibility of CCTV Image Evidence", Journal of Digital Forensics, vol. 9(1), pp. 47-67. 2015.
- [3] J. Ruiz-Santaquiteria, A. Velasco-Mata, N. Vallez, G. Bueno, J. A. Álvarez-García and O. Deniz, "Handgun Detection Using Combined Human Pose and Weapon Appearance", IEEE Access, vol. 9, pp. 123815-123826, 2021.
- [4] 법률 제 16930 호 개인정보보호법, 제 25 조 2 항. 2020.
- [5] Frederique Crete, Thierry Dolmiere, Patricia Ladret, Marina Nicolas, "The blur effect: perception and estimation with a new no-reference perceptual blur metric", Human Vision and Electronic Imaging XII, 64920I, 12 February 2007.
- [6] G. M. Faustino and L. H. de Figueiredo, "Simple Adaptive Mosaic Effects", XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'05), pp. 315-322, 2005.
- [7] H. Fang, S. Xie, Y. Tai and C. Lu, "RMPE: Regional Multi-person Pose Estimation", IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2353-2362, 2017.
- [8] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. Gehler, and B. Schiele, "Deepcut: Joint subset partition and labeling for multi person pose estimation.", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.
- [9] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model", European Conference on Computer Vision (ECCV), May 2016.
- [10] E. Insafutdinov, M. Andriluka, L. Pishchulin, S. Tang, E. Levinkov, B. Andres, and B. Schiele, "Arttrack: Articulated multi-person tracking in the wild", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.