

지능형 영상 감시 시스템 기반의 실시간 아동학대 발생 보고 시스템

구영준*, 김동욱*, 윤서원*, 경예지*, 성경화*, 고석주*, 김정석**

*경북대학교 컴퓨터학부

**SK 텔레콤

e-mail : guyeongjun@naver.com, gygacpu@naver.com, floidea@naver.com, yg9353@gmail.com,
dwg04031@gmail.com, sjkoh@knu.ac.kr, justin.joy.9to5@gmail.com

A Study on the Real-Time Child Abuse Reporting System based on Intelligence Video Surveillance System

Youngjun Gu*, Dongwook Kim*, Seowon Yun*, Yeji Kyeong*, Gyeonghwa Sung*, Seokjoo Koh*,
Jeongseok Kim**

*Kyungpook National University, Computer Science and Engineering

**SK Telecom

요 약

아동 교육기관 내에서 발생하는 아동학대 빈도가 높아짐에 따라 빠른 대처의 필요성이 증가하고 있다. 본 논문은 영상 분석을 통해 실시간으로 특정 상황 발생을 인식하고 자동으로 대응 가능한 지능형 영상 감시 시스템을 적용하여 학대 발생 시 즉각적인 조치를 하고 추후 아동학대 발생을 방지할 수 있는 시스템을 제안한다. 본 논문의 구성은 다음과 같다. 1장 서론에서 시스템의 제안 배경과 필요성을 설명하고, 2장 본론에서는 시스템의 전반적인 구조를 소개한다. 또한, 각각의 서버와 프로세스의 기능과 구조, 동작에 대해 설명한다. 3장 결론에서는 연구 결과로 파악한 시스템의 아키텍처를 제안한다.

1. 서론

우리는 현재 CCTV의 홍수 속에 살고 있다. ‘국가지표체계’에 따르면 2017년 전국에 설치된 공공 CCTV는 약 95만 대이며, 전년 대비 12.9% 증가했다.[1] 건물 내에 설치된 민간 CCTV까지 합산하면 그 수는 약 150만 대로 추정된다. 범죄를 비롯한 비상 상황을 예방하기 위해 CCTV를 많이 설치하는 추세이지만 CCTV는 수집한 영상을 지속적으로 저장할 뿐 특정 상황이 발생했을 때 그에 따른 조치를 취하는 기능은 없다. 따라서 실시간 모니터링을 통해 범죄 등의 특정 상황이 발생하는지 직접 확인해야 한다. 이에 따라 영상 분석을 통해 실시간으로 특정 상황의 발생을 인식하고 자동으로 대응할 수 있는 지능형 영상 감시 시스템에 대한 관심도가 높아지고 있다.

어린이집 아동학대 발생 현황

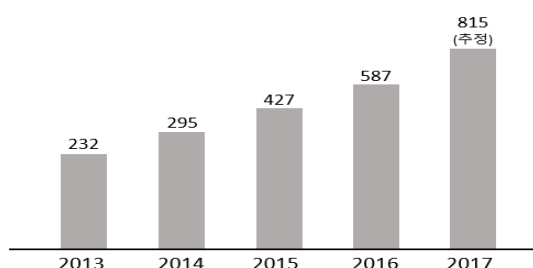


그림 1 어린이집 아동학대 발생 현황

그림 1과 같이 요즘 유치원, 어린이집 등 아동 교육기관 내에서 발생하는 학대의 빈도가 점점 늘어나고 있으며 학대로 인해 아동이 사망하는 경우도 종종 발생하여 심각한 사회적 문제로 대두되고 있다. 또한 최근 헌법재판소는 어린이집 내 CCTV 설치 의무화가 합헌이라는 결정을 내렸다.[2] 이를 통해 국가 차원에서도 아동학대 발생의 방지를 위해 많은 노력을 기울이고 있음을 알 수 있다. 하지만 앞서 얘기했듯이 CCTV는 실시간 모니터링을 필요로 하며 이로 인해 상황 발생 후 대처까지 걸리는 시간이 비교적 길다.

따라서 본 고에서는 지능형 영상 감시 시스템을 적용하여 아동 교육기관 내에서 학대 발생 시 피해 아동의 부모에게 즉시 알려주는 등 즉각적인 조치를 가능하게 하고 추후 아동 학대 발생을 방지할 수 있는 보다 체계적인 시스템을 제안한다.

2. 본론

2.1 시스템 구조

전체 시스템 구조는 그림 2와 같다. 먼저 CCTV는 실시간으로 스트리밍 서버에 영상을 전송한다. 스트리밍 서버는 CCTV로부터 수신한 영상을 분석 서버의 폭력 행위 감지 프로세스로 전송한다. 이때, 스트리밍 서버는 실시간 스트리밍을 위해 해당 기능 구현이 가능한 오픈소스 멀티 미디어 프레임워크인

GStreamer 를 사용한다. 분석 서버의 폭력 행위 인식 프로세스는 스트리밍 서버로부터 수신한 영상 내의 폭력 상황 발생 여부를 확인한다. 만약 폭력 상황이 발생했다면 해당 영상을 얼굴 인식 프로세스로 전송한다. 얼굴 인식 프로세스는 폭력 행위가 감지된 영상을 프레임 단위로 잘라 한 프레임마다 얼굴 인식 모델을 적용하여 피해 원아를 식별할 때까지 반복된다. 이후 데이터베이스에 저장된 원아 정보에 상응하는 보호자 정보를 얻어 해당 피해 원아의 보호자에게 학대 상황이 발생했음을 실시간으로 보고한다.

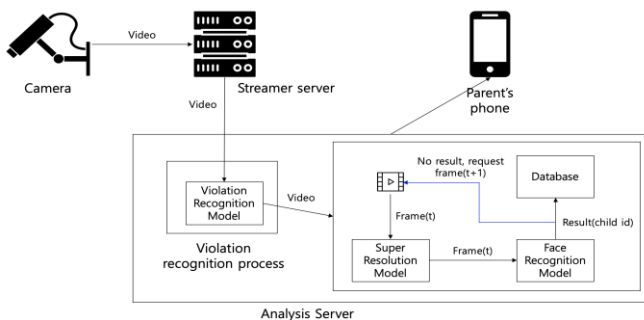


그림 2 시스템 구성도

2.2 실시간 비디오 스트리밍 서버

해당 시스템은 CCTV 와 같은 디바이스에서 수집한 영상을 스트리밍하기 위해 GStreamer 를 사용한다. GStreamer 는 오디오, 비디오에 상관없이 모든 종류의 멀티 미디어 데이터 흐름을 처리할 수 있으며, MP2, Ogg/Vorbis, MPEG-1/2, AVI, Quicktime, mod 등과 같은 다양한 포맷을 지원하는 미디어 플레이어 생성 컴포넌트를 포함하고 있다. 따라서 수집한 영상을 적절한 형식의 데이터로 실시간 스트리밍하여 퍼블릭 클라우드 서버로 전송할 수 있다.

퍼블릭 클라우드가 제공하는 데이터 스트림을 통해 대규모 레코드 스트림을 실시간으로 수집하고 처리할 수 있다. 데이터 스트림에 수집된 미디어 파일 형태로 저장소에 저장, 관리가 가능하고 일정 기간 데이터를 보존할 수 있다. 또한, 클라우드를 통해 실시간으로 대량의 미디어 파일을 분석하기 위한 다중 레코드 처리가 가능하며 GStreamer 로부터 받은 스트리밍 데이터를 분석할 수 있는 스트림 프로세스(Stream Processor)도 제공한다. 그리고 API 가 미디어 파일을 분석하기 위해 데이터 스트림으로 전송하는 분석 결과를 읽을 프로세서도 사용할 수 있다.

따라서 클라우드가 제공하는 API 를 사용하여 저장소에 저장한 미디어 파일들 중에서 선택한 Input 미디어 파일의 인물 존재 여부를 확인하고 존재하지 않는 경우, 폭력 감지의 필요성이 없다고 판단하여 해당 미디어 파일은 폭력 행위 감지 프로세스 진행 서버로 전송하지 않는다. 인물이 존재할 경우 해당 미디어 파일을 서버로 전송한다.

2.3 폭력 행위 감지 프로세스

카메라 영역 내에서 인물을 감지할 경우 스트리밍 서버로부터 영상을 수신한다. 이후 폭력 감지를 위한 딥러닝 모델[3]을 적용하여 해당 영상에서 폭력 행위가 검출되는지 확인한다. 이때 사용하는 딥러닝 모델은 영상 내의 연속한 두 프레임을 입력 데이터로 하며 두 프레임 간의 움직임 사이 발생하는 차이를 비교하고 이 움직임이 폭력 행위인지 파악한다. 영상에서 폭력 행위가 검출될 경우 피해 원아의 정보를 알기 위해 얼굴 인식 프로세스를 실시한다.

2.4 얼굴 인식 프로세스

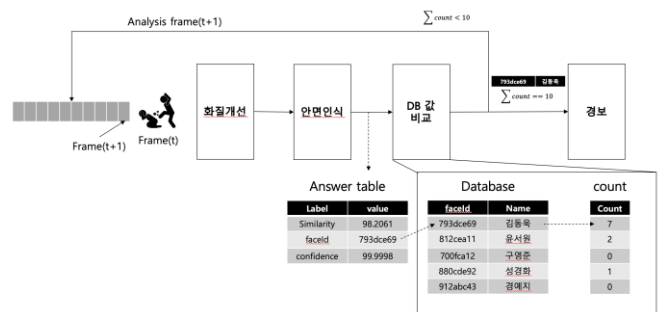


그림 3 안면인식 과정

2.4.1 화질 개선 모델

CCTV 의 화질이 좋지 않을 경우 얼굴 인식 모델의 정확도가 높다고 하더라도 인식률이 낮아질 수 있다. 해당 문제점을 고려하여 본 시스템은 얼굴 인식 모델을 적용하기 전에 화질 개선 모델을 통해 저화질의 영상 프레임을 고화질로 개선하는 과정을 진행한다. 결과로 얻은 고화질의 영상 프레임에 얼굴 인식 모델을 적용하여 해당 원아를 식별한다.

2.4.2 얼굴 인식 모델

해당 모델은 퍼블릭 클라우드가 제공하는 API 를 사용하며, 화질 개선 모델을 통해 얻은 고화질 이미지 프레임을 Input Data 로 하여 얼굴 인식을 진행한다. 이 과정은 다음과 같다. 먼저 탐색 대상인 원아들의 이미지를 바탕으로 얼굴 모음을 구성한다. 이후 Input Data 가 들어오면 얼굴 모음에서 일치하는 얼굴이 존재하는지 확인하고, 존재하는 경우 해당 얼굴의 Face-ID 와 인식률, 정확도, 얼굴 정보 등의 데이터를 반환한다.

이때, 얼굴 인식의 정확도 향상을 위해 반환된 정확도가 80% 미만일 시 데이터를 저장하지 않고 다음 프레임의 화질 개선을 진행하며, 정확도가 80% 이상이면 해당 프레임의 정보를 저장한다. 저장된 프레임의 정보가 10 개가 되면 최빈값인 Face-ID 를 선택하여 시스템 서버로 전송한다. 서버는 해당 Face-ID 를 보유한 원아의 보호자 연락처를 데이터베이스에서 찾아 학대가 발생했음을 알린다. 그림 3 은 화질개선부터 알림 서비스까지 전체적인 안면 인식 과정을 도식화하였다.

2.4.3 데이터베이스 구축

원아 데이터베이스는 얼굴 인식 모델에서 받은 Face-ID 를 Primary Key 로 하는 Child 테이블로 구성한다. Child 테이블의 칼럼은 Face-ID(Primary Key), 원아 이름, 나이, 성별, 보호자 연락처로 구성한다.

3. 결론

본 고에서는 CCTV로부터 수집한 영상 분석을 통해 폭력 행위를 검출하고 이때 피해 아동을 식별하기 위해 얼굴 인식 모델을 적용하여 해당 보호자에게 실시간으로 학대 상황 발생을 보고하는 지능형 영상 감시 시스템 기반의 실시간 아동학대 발생 보고 시스템에 대한 전반적인 구조와 프로세스에 대해 설명하였다.

오픈소스 멀티 미디어 프레임워크인 Gstreamer 를 이용하여 비디오 데이터를 처리할 경우 클라우드 서버로 실시간 스트리밍이 가능하며, 클라우드가 제공하는 API 를 통해 이후 프로세스의 진행 여부 판단이 가능함을 확인하였다. 또한, 딥러닝 모델을 적용하여 폭력 행위를 검출한 후 얼굴 인식 프로세스를 진행할 때 어떤 종류의 데이터를 반환하는지 확인하였으며, 결과 데이터를 바탕으로 원아 정보에 관한 데이터베이스 구축과 이를 이용한 알림 서비스 구현이 가능함을 확인하였다.

향후 연구 결과를 바탕으로 실제 시스템을 구현할 예정이다. 본 시스템은 심각한 사회적 문제로 여겨지는 아동학대의 발생에 대한 즉각적인 대처를 가능하게 하여 추후 교육기관 내 아동학대 발생을 방지하는데 긍정적인 영향을 끼칠 것이다.

Acknowledgement

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음(2015-0-00912)

This research was supported by the Korean MSIT (Ministry of Science and ICT), under the National Program for Excellence in SW supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation)(2015-0-00912)

참고문헌

- [1]E-나라지표, “공공기관 CCTV 설치 및 운영현황”
http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=2855, 2019.04.10
- [2]임현주, “현재 “어린이집 CCTV 의무화 합헌” ... 아동학대 방지”, MBC 뉴스, 2018.01.01 작성
http://imnews.imbc.com/replay/2018/nw1800/article/4487840_22625.html
- [3] Violence Detection by CNN + LSTM
<https://github.com/JoshuaPiinRueyPan/ViolenceDetection>