이상행동 감지 모델 시스템

김승준[O,1], 우승택[1], 박한솔[1], 김세아[1], 김민준[1], 정설영[1], 김호영[2]

경북대학교 컴퓨터학부[1], (주)우경 정보 기술[2]

kimmokalover@gmail.com, dntmdxor99@naver.com, pjhcsols@naver.com,
seakim@knu.ac.kr, kmjj5275@naver.com, snowflower@knu.ac.kr,
hykim@wkit.co.kr

System for Anomalous Behavior Detection

SeungJun Kim[O,1], SeungTaek Woo[1], HanSol Park[1], SeAh Kim[1], MinJun Kim[1], SeolYoung Jeong[1], HoYoung Kim[2]

School of Computer Science and Engineering, Kyungpook National University[1],
Wookyung Information Technology[2]

요 약

최근 "묻지마 범죄"가 증가하고 있어 이러한 범죄를 줄이기 위한 노력이 확대되고 있다. 현재 이상행동 감지 시스템은 두 가지 주요 방법을 사용하고 있다. 직접 관제인에 의한 수동 감시 방식과 인공지능을 활용한 이상행동 감지 방식이다. 직접 관제인을 통한 감시는 인력 소모가 크다는 단점이 있다. 하지만 AI 기반의 이상행동 감지는 일반적으로 공간 및 시간적 특성을 구별하여 객체와 그들의 움직임 정보를 추출하고, 공간적 및 시간적 특성을 결합하여 객체의 신원과 행동을 분류하기 때문에 인력 소모나, 경제적으로 효과적이다. 하지만 대부분의 인공지능 모델은 현실의 비디오의 상태를 고려하지 않고, 비디오의 모든 프레임을 동일하다고 가정하고 있다.

그러나 Low-Delay 모드와 같이 I-Frame과 P-Frame이 있는 경우, P-Frame은 일반적으로 I-Frame에 비해 정보가 적을 수 있으므로 이러한 가정은 모델이 학습을 불안전하게 할 가능성이 있다. 또한 QP(Quantization Parameter)가 높을수록 비디오 프레임의 상세 정보가 크게 손실되어 모델이 움직임과 행동을 정확하게 파악하는 데 어려움을 겪을 수 있다. 따라서 본 논문에서는 비디오 부호화 기술을 고려하여 인공지능 모델을 사용한 웹 기반 감지 시스템을 제안한다. 구체적으로는 인공지능 모델을 사용하여 관리자에게 실시간으로 알림을 보내고 해당 내용을 다중 서버 구조를 통해 처리해, 사용자가 웹 기반 UI를 통해 이상행동을 확인할 수 있는 시스템 모델을 설계하고자 한다.

1. 서 론

범죄는 사회 안전과 안정에 심각한 영향을 미치는 사회 문제 중 하나로, 이를 예방하고 대응하기 위한 다양한 노력이 이루어지고 있다. 이러한 노력 중 하나는 현대 기술과 인공지능의 활용하여 안면 인식, 프레드폴 알고리즘 등이 있다. 본 논문에서는 비디오 객체 행동 인식 시스템을 통해 범죄와 연관된 문제를 다루고자 한다.

비디오 객체 행동 인식은 객체의 움직임과 행동을 이해하고 분류하는 기술로, 이를 통해 범죄 예방 및 수사에 도움을 줄 수 있다. 현재, 딥러닝과 비디오 부호화 기술, 그리고 다양한 웹 프레임워크의 발전으로 비디오 객체 행동 인식 모델의 정확도와 성능이 크게 향상되고, 사용자 친화적인 시스템들이 만들어지고 있다. 그러나 기존의 인공지능 모델들은 비디오의 상태를 고려하지 않고, 모든 프레임을 동일하게 다룬다. 하지만 대부분 현실의 비디오는 압축된 상태로, 현재 많이 사용되는 기술인 H.264/AVC(Advanced Video Coding)와 H.265/HEVC(High Eifficiency Video Coding)로 압축되어 있어, Low-Delay와 Random-Access 모드 과정에서 프레임 정보 손실과 화질 변화가 발생하기 때문이다.[1][2] 이로 인해 객체의 움직임과 행동을 정확하게 파악하기 어려워진다. 그리고 기존의 관제 시스템은 실시간으로 동영상 데이터를 처리하는 특성 때문에 서버의 오버헤드가 크다.

본 논문에서는 이러한 관제 시스템의 특성과 비디오 부호화 기술의 특성을 고려하여, 인공지능 모델의 성능을 향상하고 효율적으로 영상 정보를 처리할 수 있는 자동 관제 시스템을 제안한다. 본 시스템은 기존의 인공지능 기반 행동 분석 모델을 새롭게 만드는 것이 아니라, 입력 데이터의 특성을 고려하여 모델의 성능을 향상하는 데 중점을 두고 있다. 이는 범죄와 관련된 객체 행동 인식 분야에 활용될 수 있으며, 추후 범죄 예방 및 수사에 적극 활용될 수 있다. 그리고 실시간 범죄예방을 위한 실시간 정보 처리와 이를 통해 범죄와 관련된 비디오 객체 행동 인식 분야에서의 연구와 응용에 새로운 가능성을 제시한다.

2. 관련 연구

2.1 비디오 행동 인식

SlowFast는 비디오 행동 인식 딥러닝 모델이다.[4] 해당 모델은 공간적 구조를 파악하는 Slow Stream과 시간적 특징을 파악하는 Fast Stream을 분리하여 학습한다.

해당 모델의 Slow Stream에서는 모델에 입력되는 프레임의 시퀀스에서 τ만큼의 프레임 간격을 두어 하나의 프레임을 선택해 일정한 개수의 프레임 스택을 Stream에 입력한다. τ는 일반적으로 16으로 설정한다. 이는 30fps의 비디오에서 초당 약 2개의 프레임을 얻는 것과 같다. 따라서 Slow Stream에서는 시간적인 특징보다 공간적인 구조에 집중하고, 프레임의 객체가 무엇인지에 대한 정보를 학습한다.

Slow Stream과 반대로 Fast Stream에서는 시간적인 정보를 학습한다. 이때 Slow Stream보다 작은 프레임 간격을 두어 시간적인 특성을 파악한다. 이때 Fast Stream이 사용하는 간격은 τ/α이다. 일반적으로 α는 8로 설정한다. 따라서 τ가 16이고, α가 8이라면 Slow Stream보다 8배 많은 프레임 스택을 얻게 된다. 이는 30fps 비디오에서 초당 약 15개의 프레임을 얻는 것과 같다.

마지막으로 Fast Stream은 Slow Stream에 비해 적은 채널 용량을 가지고 있다. Slow Stream의 채널 용량이 C일 때, Fast Stream의 채널 용량은 C*β로 설정한다. 이때 β는 일반적으로 1/8로 설정한다. Fast Stream이 적은 채널 용량을 가진다는 것은 공간적인 의미를 파악하는 능력이 Slow Stream에 비해 약하다는 것을 의미하며, 이는 Fast Stream이 시간적 특징에 집중하도록 만든다.

그림 1은 해당 모델의 개요이다. 그림 1에서 볼 수 있듯이 단계마다 Fast Stream의 정보를 Slow Stream으로 입력한다. 최종적으로 해당 모델은 Fast Stream과 Slow Stream의 정보를 종합하여 비디오의 객체가 어떤 행동을 하고 있는지 학습하게 된다.

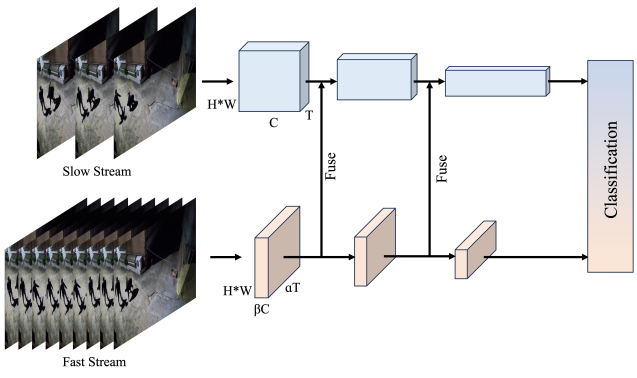

그림 1. SlowFast 모델 개요

2.2 분산 서버 구조

다중 서버 구조란 두 개 이상의 서버를 연계, 운용하는 기술로서 여러 개의 Transaction 요청이 들어왔을 때, 급격한 트래픽 변화에 대응하고 안정적으로 서비스를 제공하기 위한 기술이다. 분산 서버 처리를 위한 기술로는 대표적으로 로드밸런싱이 있는데, 이는 외부의 요청을 서버에 적절히 분산하여 처리하는 것을 의미한다.

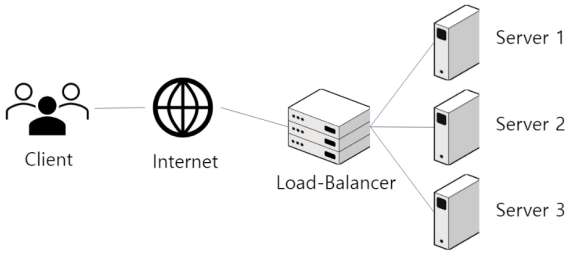

그림 2. 로드 밸런싱 개요

2.3 비디오 부호화 기술

2.3.1 Low-Delay 모드

Low-Delay 모드는 화상 통화, 스트리밍 방송, CCTV 등 다양한 분야에서 사용되는 비디오 부호화 모드이다. 해당 모드는 과거의 프레임을 참조하는 방식인 Inter 프레임 코딩을 사용하여 부호화하는 P-Frame과 프레임 내부의 정보만을 활용하는 방식인 Intra 프레임 코딩을 사용하여 부호화하는 I-Frame으로 이루어진다.

그림 3에서 볼 수 있듯이 하나의 GOP 내에는 하나의 I-Frame과 다수의 P-Frame이 존재한다. 이때 I-Frame을 부호화하는 양자화 파라미터인 QP는 일반적으로 가장 낮은 값을 가진다. 따라서 I-Frame은 가장 좋은 화질을 갖게 되지만 압축 효율이 감소한다. P-Frame을 부호화하는 QP는 일반적으로 I-Frame보다 높은 QP를 가지지만, 중간 혹은 끝부분에 I-Frame과 같은 QP를 가질 수도 있다. 결론적으로 P-Frame은 I-Frame에 비해 화질이 낮아지고, 세부 정보가 손실된다.

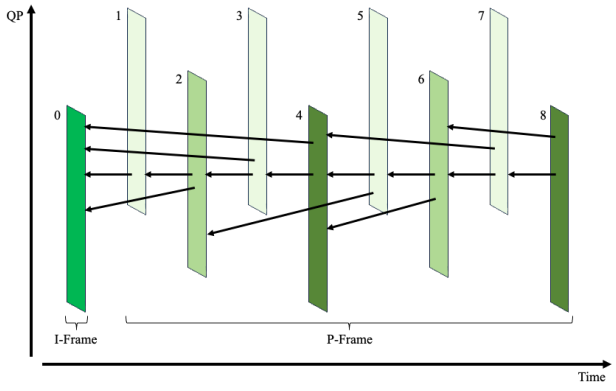

그림 3. Low-Delay 모드

2.3.2 Random-Access 모드

Random-Access 모드는 고화질 동영상 서비스를 제공하는 경우 등 다양한 분야에서 사용한다. 일부 지연을 허용하면서 우수한 화질을 제공하며, 임의접근을 지원한다. 해당 모드는 Low-Delay와 다르게 과거와 미래 프레

임까지 활용하는 B-frame을 추가로 사용한다. 그림 4에서 볼 수 있듯이 Random-Access 모드에서는 부호화 순서와 디스플레이 순서가 다를 수 있다. 또한, 임의접근을 지원하기 위해 주기적으로 I-Frame을 삽입한다. 이때 B-frame 과 P-frame은 I-frame보다 QP가 높거나 프레임 코딩 방식의 특징 때문에 화질이 낮고, 세부 정보가 손실될 수 있다.

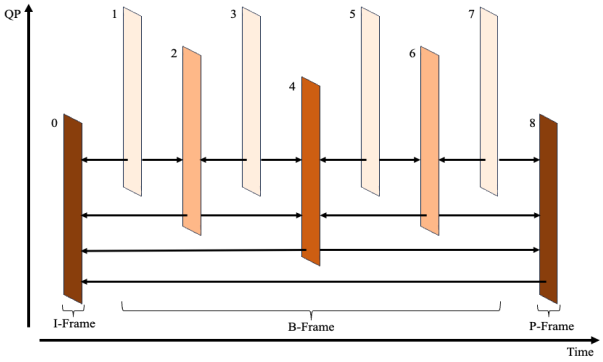

그림 4. Random-Access 모드

2.4 관제 시스템 아키텍쳐

그림 5처럼 기존의 관제 시스템은 일반적으로 중앙화된 아키텍처를 사용한다. 이는 영상 처리와 사용자 요청 처리를 하나의 중앙 서버에서 수행하는 방식을 의미한다. 이러한 아키텍처는 초기에 구현하기 쉽고 간단한 시스템을 구축하는 데 도움이 될 수 있으나, 여러 단점이 있다.

먼저 중앙 서버가 모든 작업을 처리하기 때문에 시스템의 확장성이 제한된다. 증가하는 트래픽 또는 요청에 대응하기 위해 중앙 서버를 계속 강화하거나 대규모 서버를 사용해야 할 수 있다. 따라서 비용이 증가하고, 과도한 자원 사용과 함께 단일 지점 장애의 위험이 증가한다. 또한 중앙 서버가 모든 요청을 처리하기 때문에 성능 병목 현상이 발생할 수 있으며, 서버 과부하로 인한 응답 지연 문제가 발생할 수 있다.

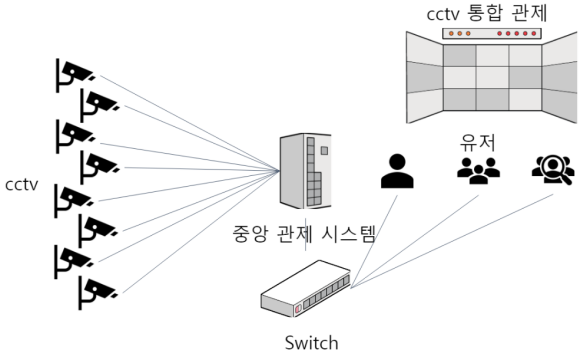

그림 5. 기존 관제 시스템 아키텍처

3. 시스템 설계

3.1 제안 인공지능 모델

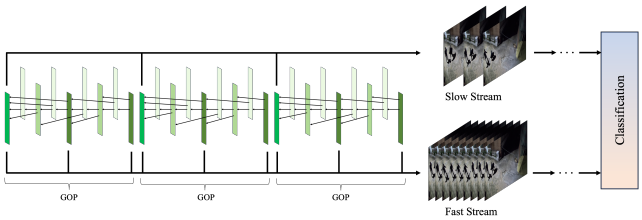

그림 6. 제안 인공지능 모델

앞선 기존 SlowFast 모델은 비디오 내부의 프레임 간의 특성을 고려하지 않고, 일정 구간마다의 프레임을 바탕으로 학습한다. 만약 Low-Delay로 부호화 복호화 과정을 거친 비디오를 학습할 때, Slow Stream에 입력되는 프레임이 모두 높은 QP를 가지는 P-Frame이라면 Slow Stream은 상대적으로 정보가 손실된 프레임만 선택하게 된다. 이러한 과정을 통해 학습된 모델은 최적화되지 못하고, 추론 과정에서 원하는 결과가 나오지 못할 가능성이 존재한다. 또한 Fast Stream에서도 높은 QP를 가지는 P-Frame 위주로 선택한다면, 프레임 내의 노이즈, 모션 블러로 인해 객체의 행동을 정확히 표현하지 못할 여지가 있다.

Random-Access 모드 또한 Low-Delay와 같이 프레임마다 특성이 다르고 정보의 양이 다르다. 해당 경우 또한 모델이 학습할 때, GOP 내에서 정보의 양이 상대적으로 적거나, 화질이 좋지 않은 프레임을 바탕으로 학습한다면, 비디오 내부의 정보를 완전히 활용하지 못한 학습 결과가 될 수 있다.

Low-Delay 모드와 Random-Access 모드에서 GOP 내에 정보 손실이 적은 프레임과 큰 프레임이 존재한다. 앞선 문제를 해결하기 위해 그림 6처럼 Slow Stream에 정보 손실이 적은 프레임인 I-Frame 혹은 QP가 낮은 프레임들로만 넣게 된다면, 모델은 프레임 내의 객체를 명확히 인식할 수 있으며, 이를 통해 최적화된 결과에 도달할 수 있다. 또한 Fast Stream에도 정보 손실이 적은 프레임 위주로 학습한다면, 객체의 행동을 보다 잘 파악할 수 있고, 이를 통해 성능이 향상된다.

3.2 제안 시스템 아키텍처

기존의 아키텍처는 과도하게 중앙화 되어있어, 급격한 트래픽 변화에 대응하기 쉽지 않다. 또한 관제 시스템의 특성인 멀티미디어 전송은 일반적인 Transaction 처리에 비해 큰 비용을 요구하기 때문에 사용자에게 버퍼링과 같은 문제가 발생할 수 있다. 따라서 기존 중앙 집중적 서버의 단점을 극복한 다중 서버 기반 관제 시스템을 제안한다.

제안하는 시스템의 구조는 그림 7이다. CCTV를 통해 생성된 영상들은 비디오 제어 서버를 통해 기록되고 동시에 인공지능 서버를 통해 영상이 실시간으로 전달된다. 인공지능 서버에서는 실시간으로 영상 내에 이상행동 객체가 존재하는지 판단하고, 존재할 경우 해당 프레임과 메타 데이터를 클라우드 기반 파일 시스템 서버에 전달해 저장한다.

클라이언트 서버가 실시간 혹은 과거 영상 요청을 한

다면 로드밸런서를 통해 가장 오버헤드가 적은 서버의 주소를 알려주고, 클라이언트 서버는 해당 주소를 통해 파이프라인을 만들어서 전용 점대점 연결을 한다. 이때 사용되는 Scheduling 기법은 프로세스 수를 고려하는 Multi-Level Queue Scheduling을 사용한다.

파이프라인이 구축되면, 클라이언트가 연결을 종료할 때까지 계속해서 멀티미디어 데이터를 전송한다. 그리고 데이터가 저장된 클라우드에 접근하는 상황에도 클라우드 로드밸런서를 사용하여 I/O 연산량이 가장 적은 클라우드와 연결한다. 이를 통해 멀티미디어 전송 오버헤드를 최대한으로 줄임으로써 실시간성을 극대화하고, 사용자 친화적인 서비스를 제공할 수 있다.

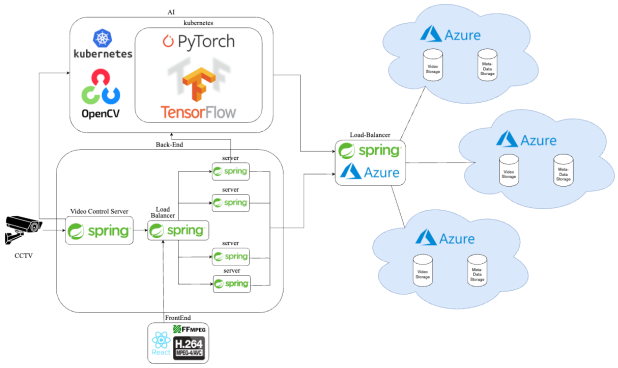

그림 7. 제안 시스템 아키텍처

4. 결론

기존의 관제 시스템은 중앙 집중적인 서버 구조로 과도한 트래픽 상황에서 실시간으로 동영상과 같은 멀티미디어를 전송할 때 문제가 발생할 수 있다. 또한 기존의 비디오 행동 인식 모델은 대부분 모델의 구조를 변경하거나, 더 깊고 넓은 네트워크를 구성하여 시스템의 성능을 향상하려 한다. 하지만 이러한 방법은 비디오의 모든 프레임이 압축되지 않고, 프레임마다 같은 정보의 양을 가지고 있음을 가정한다.

본 논문에서는 분산 서버 기반의 관제 시스템 모델과 함께 객체의 이상행동을 기존 인공지능 모델보다 더 정확하게 판별할 수 있는 시스템을 제안한다. 비디오 부호화 기술의 특성을 파악한 후, 이를 명시적으로 활용하는 모델을 같이 사용하여 미래에 발생할 수 있는 범죄를 효과적으로 예방할 수 있다.

5. Acknowledgement

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음" (2021-0-01082)

6. 참고 문헌

[1] Tamhankar, A., & Rao, K. R. (2003, July). An overview of H. 264/MPEG-4 Part 10. In*Proceedings EC-VIP-MC 2003. 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications (IEEE Cat. No. 03EX667)*(Vol. 1, pp. 1-51). IEEE.

[2] Sullivan, G. J., Ohm, J. R., Han, W. J., & Wiegand, T. (2012). Overview of the high efficiency video coding (HEVC) standard.*IEEE Transactions on circuits and systems for video technology*,*22*(12), 1649-1668.

[3] Feichtenhofer, C., Pinz, A., & Zisserman, A. (2016). Convolutional two-stream network fusion for video action recognition. In*Proceedings of the IEEE conference on computer vision and pattern recognition*(pp. 1933-1941).

[4] Feichtenhofer, C., Fan, H., Malik, J., & He, K. (2019). Slowfast networks for video recognition. In*Proceedings of the IEEE/CVF international conference on computer vision*(pp. 6202-6211).