

실시간 행동 감지 기술을 이용한 행동 수집 서비스

CCTV Service Using Real-time Action Detection

김동진, 오병현, 윤태성
건국대학교 컴퓨터공학부

kdje0506@konkuk.ac.kr, hhyeon55@konkuk.ac.kr, greatstary@konkuk.ac.kr

연구 배경 및 목적

[연구 배경]

- 버스 기사 폭행, 어린이 집 유아 폭행과 같은 폭행 범죄가 자주 발생
- 실생활에서 CCTV가 사용되고 있으나, 사람이 직접 영상을 확인해야 하기 때문에 긴급 상황에 대해서는 빠른 대처가 불가능

[연구 목적]

- 기존의 CCTV 영상에 Action Detection을 수행하여 특정 행동을 감지함으로써 긴급 상황이 발생하였음을 알려주기 위함
- 영상 속 사람에 대한 행동 정보를 기록함으로써 기존의 CCTV 영상을 더 가치 있는 정형데이터로 만들기 위함

제안된 방법

1. 서비스 기능

[실시간 사람 감지 / 행동 예측]

- 촬영한 영상에 대하여 실시간으로 사람을 감지하고, 감지한 사람에 대한 행동 예측 수행

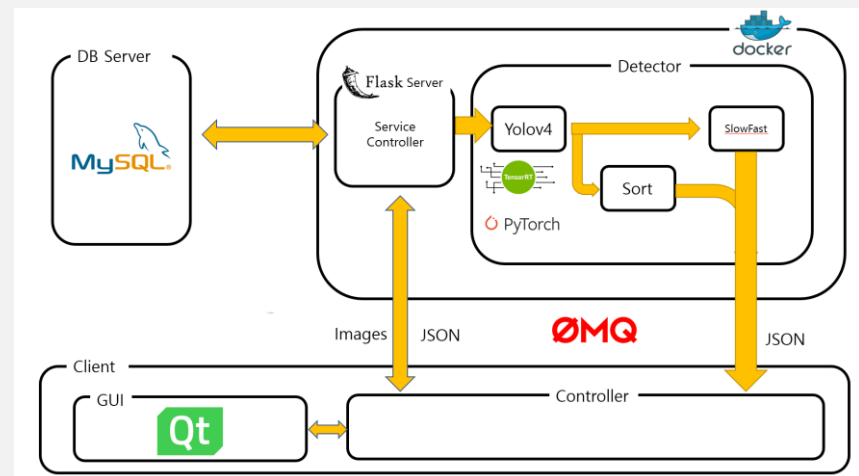
[기록 및 조회]

- 사람 위치, 예측 행동 결과는 데이터베이스에 저장
- 사용자는 검색 기능으로 결과를 조회 가능

[알림]

- 특정 행동이 발생한 경우 알림 기능

2. 모듈 설계



[그림 1] 시스템 구조도

[Client]

- PyQt를 이용하여 GUI 구성
- 사용자 인터페이스 상에서 영상파일 또는 촬영 선택 가능

[Server]

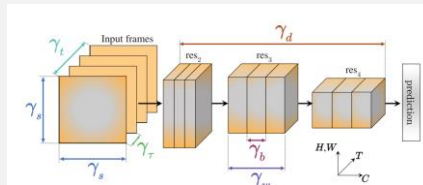
- Client와 API 통신을 위하여, Flask 을 이용,
- 성능 최적화를 위하여, multiprocessing 을 통한 병렬 처리, 비동기 통신을 이용한다.

[Action Detector]

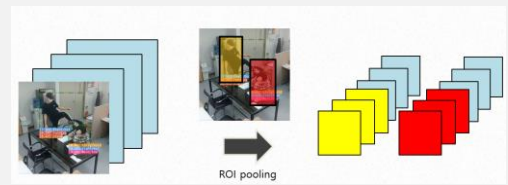
- Object Detection, Tracking Algorithm, Action Recognition으로 크게 3부분으로 모듈을 나눌 수 있다. 이때, 연산량 대비 정확도 효율에 따라 각각, YOLOv4, SORT, SlowFast(또는 X3D)을 사용

[Proposed Method]

- 이 중 X3D 모델 구조 개선을 위한 방법을 제안
- Head 계층에서 global average pooling 대신 Point-wise 3D Convolution layer 사용하여 시간 정보 손실 최소화
- ROI Pooling 된 특징과 pooling 되기 전 특징을 병합하여 context 특징 추출



[그림 2] X3D 구조



[그림 3] X3D 개선 제안

실험 결과

1. 데이터 수집

[학습 데이터]

- Pretrain: Kinetics-400
- AVA, AVA-Kinetics
- 총 100,000개 영상에서 300,000 Key Frame을 학습

[실험 데이터]

- AI HUB에서 이상행동 CCTV 영상 데이터 셋(폭행, 실신) 이용
- 직접 재현한 폭행, 실신 상황을 프로그램에 연결된 카메라로 촬영

2. 실험 결과

[동작 속도]

- 서버 사양: Intel Xeon 16 Cores 1개, Nvidia V100 1개
- Action Recognition Model으로 SlowFast R101을 사용
- 통신 지연은 고려하지 않고 측정

	FPS	SlowFast R101	X3D-M (Proposed Method)
원본 해상도	960 x 540	24.95	30.18
	1280 x 720	24.82	25.63
	1920 x 1080	21.93	22.90

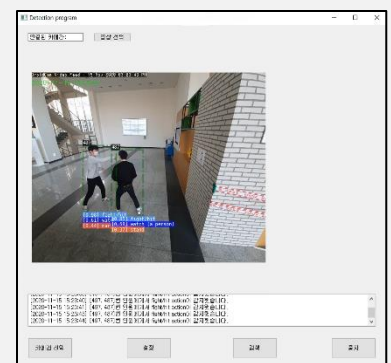
[학습 결과]

Model	Dataset	Pretrain	Val mAP	GFLOPs	Param
SlowFast R101	AVA v2.2	Kinetics-600	27.8	146.5	59.3M
X3D-M	AVA v2.2	Kinetics-600	23.2	6.2	3.1M
Proposed Model (X3D-M based)	AVA-kinetics	Kinetics-400	24.4	N/A	4.9M

[동작 사진]



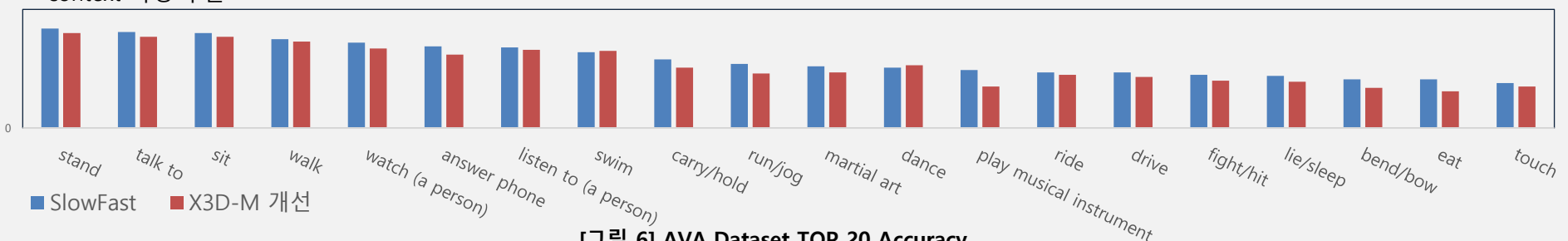
[그림 4] 실신 상황 재현



[그림 5] 폭행 상황 재현

결론

- 현재 활발히 연구되고 있는 Action Detection을 이용하여 실시간 시스템을 구현하고, 시스템 측면에서 최적화를 진행함
- 최신 모델인 X3D 구조 개선과 AVA-Kinetics 데이터셋의 학습을 통해 추론 모델 측면에서 성능 개선함
- 향후 Real-time Action Detection Service의 상용화 가능성을 확인함



[그림 6] AVA Dataset TOP 20 Accuracy