
Human Action Recognition

DMQA Open Seminar

2023.03.17

발표자 소개



❖ 안시후 (Sihu Ahn)

- Data Mining & Quality Analytics Lab (김성범 교수님)
- Ph.D. Student (2021.3 ~)

❖ 관심 연구 분야

- Computer Vision (Human Action Recognition)
- Data Science

❖ E-mail

- sihuahn@korea.ac.kr

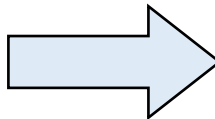
목차

1. Introduction to Human Action Recognition
2. Human Action Recognition from single-modality
 - Visual Modality
 - Non-visual Modality
3. Human Action Recognition from multi-modalities
4. Conclusion

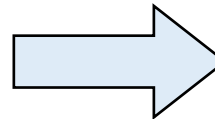
Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- 현재 행동을 인식하는 방법론
 - 걷기, 달리기 등 사람이 어떠한 행동을 하고 있는지 인식
 - 미래 행동을 예측하는 것은 다른 방법 (Human Action Prediction)



걷기



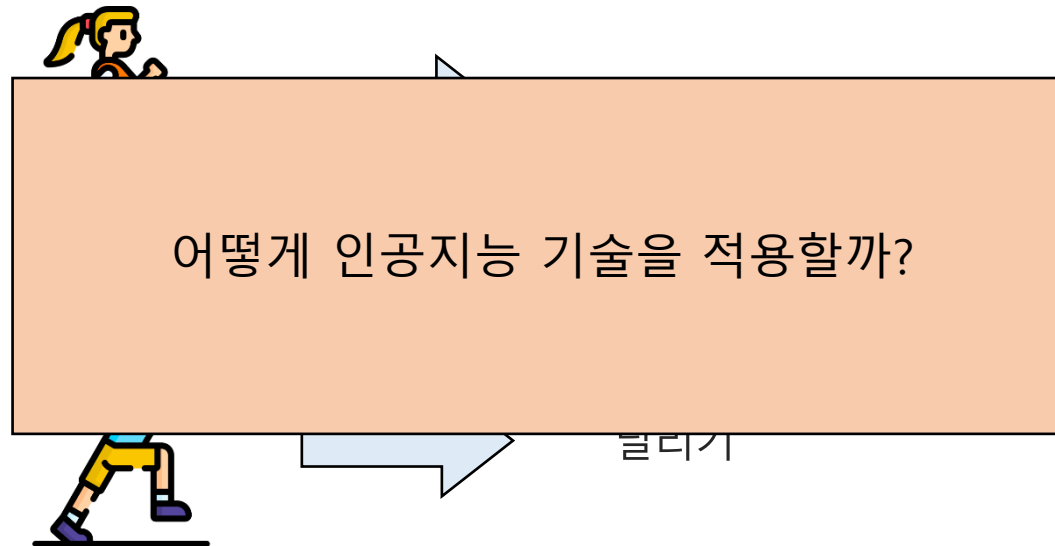
달리기

"Icon made by Freepik from www.flaticon.com"

Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- 현재 행동을 인식하는 방법론
 - 걷기, 달리기 등 사람이 어떠한 행동을 하고 있는지 인식
 - 미래 행동을 예측하는 것은 다른 방법 (Human Action Prediction)

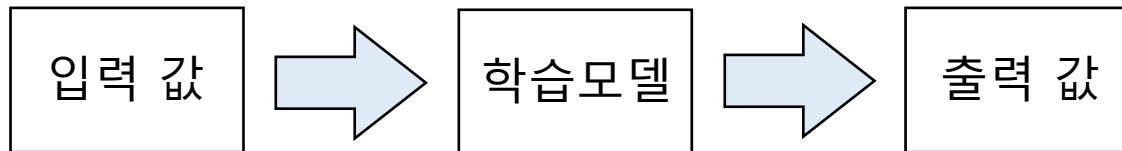


"Icon made by Freepik from www.flaticon.com"

Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

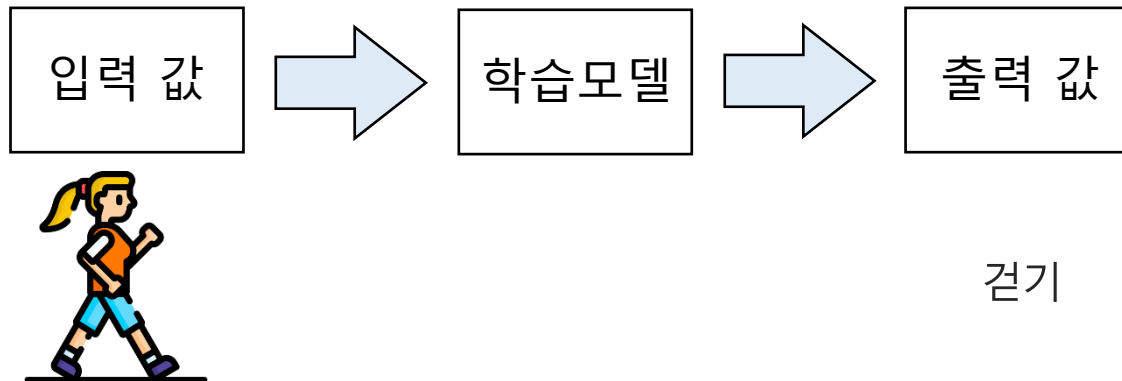
- HAR의 학습 과정
 - 인공지능 모델은 학습 모델, 입력 값, 출력 값이 필요



Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- HAR의 학습 과정
 - 인공지능 모델은 학습 모델, 입력 값, 출력 값이 필요

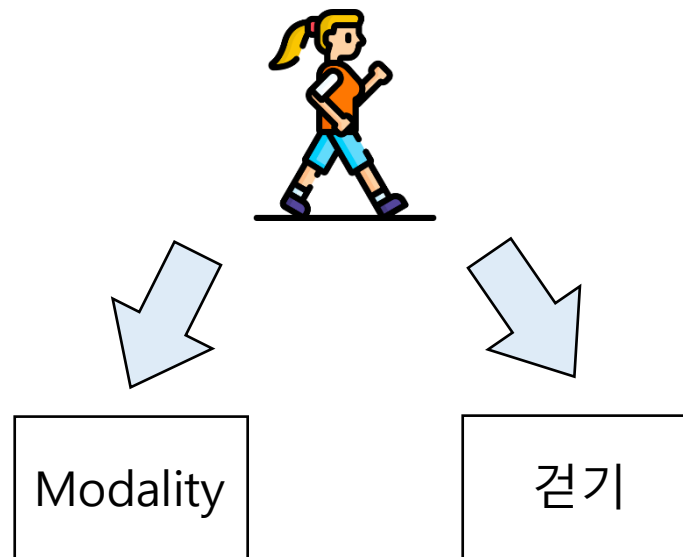


Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- HAR의 학습 과정

- 입력 값 : 현실 행동으로부터 학습에 사용 할 수 있도록 modality 데이터를 수집
- 출력 값 : 현실 행동을 이해하여 labeling 진행



"Icon made by Freepik from www.flaticon.com"

Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- HAR의 학습 과정

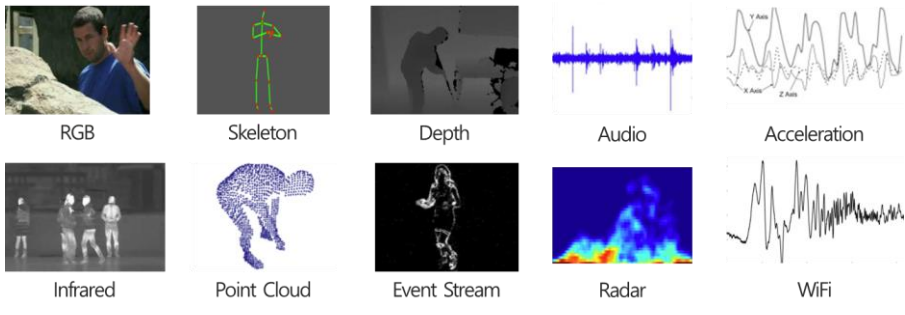
- 입력 값 : 현실 행동으로부터 학습에 사용 할 수 있도록 modality 데이터를 수집
- 출력 값 : 현실 행동을 이해하여 labeling 진행



Modality는 데이터 수집 양식을 의미



걷기

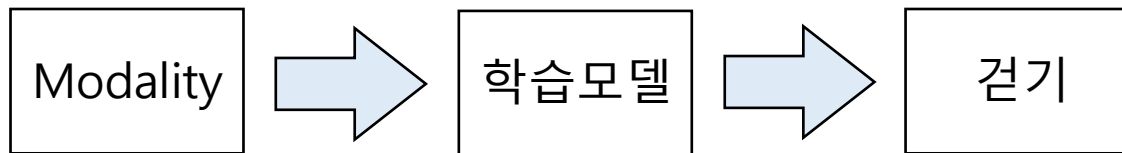


"Icon made by Freepik from www.flaticon.com"

Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- HAR의 학습 과정
 - 생성된 입력 값과 출력 값을 활용하여 학습 진행 및 행동 인식



Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- 현실에 적용된 HAR 예시
 - 현실 세상에서 범죄 예방, 게임, 스포츠 등 다양한 분야에서 활용 중



이미지 출처: 오피스 컴퍼니, 스팀, 슬릭코퍼레이션

Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- 현실에 적용된 HAR 예시 (2022 카타르 월드컵)
 - Semi Automated Offside Technology (SAOT)는 HAR 기술을 활용하여 오프사이드를 판정
 - 공에 부착된 센서와 선수를 카메라로 위치를 추적하여 오프사이드 판정에 활용

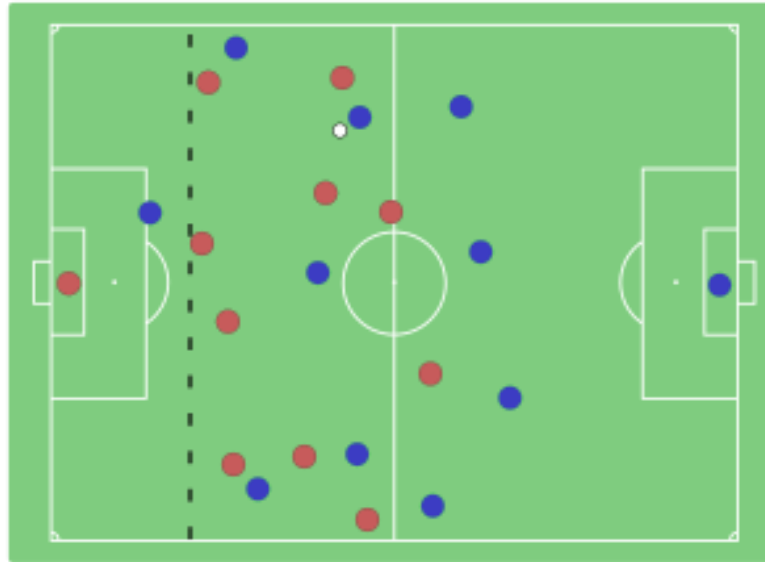


이미지 출처: FIFA

Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- 현실에 적용된 HAR 예시 (2022 카타르 월드컵)
 - 오프사이드란 축구에서 공격자 반칙 중 하나로 상대편 최후방 2번째 수비수보다 상대 골라인에 가까이 있는 경우 (오프사이드 <> 온사이드)
 - 축구계에서 가장 판정 논란이 많은 반칙

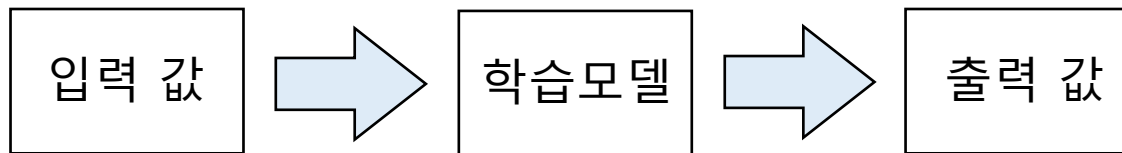


이미지 출처: 위키백과(오프사이드)

Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

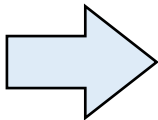
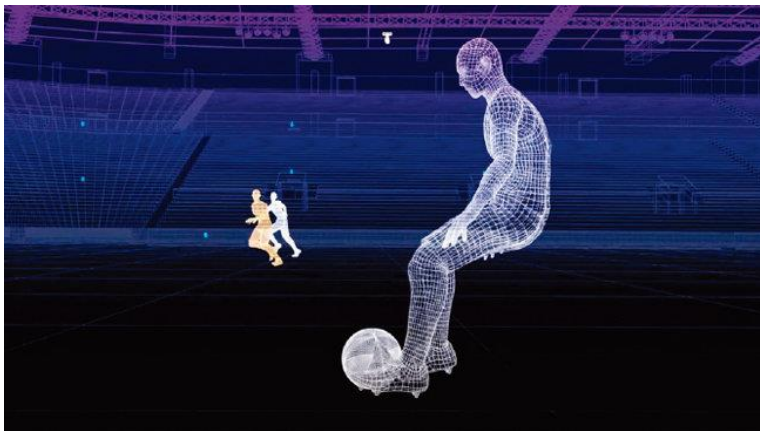
- 현실에 적용된 HAR 예시 (2022 카타르 월드컵)
 - SAOT는 HAR 기술을 활용하여 오프사이드를 판정
 - 인공지능 모델은 입력 값과 출력 값이 필요



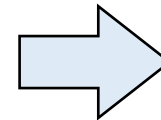
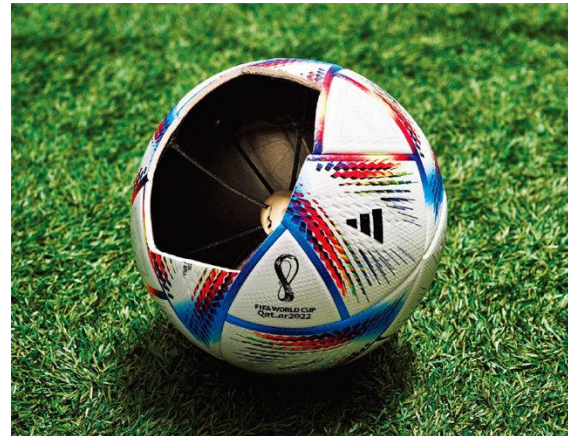
Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- 현실에 적용된 HAR 예시 (2022 카타르 월드컵)
 - 입력 값 1 : 12개의 카메라를 이용하여 사람의 위치를 추적 (50 FPS)
 - 입력 값 2 : 공에 들어있는 센서를 이용하여 위치를 추적 (500 FPS)



X, Y, Z 좌표



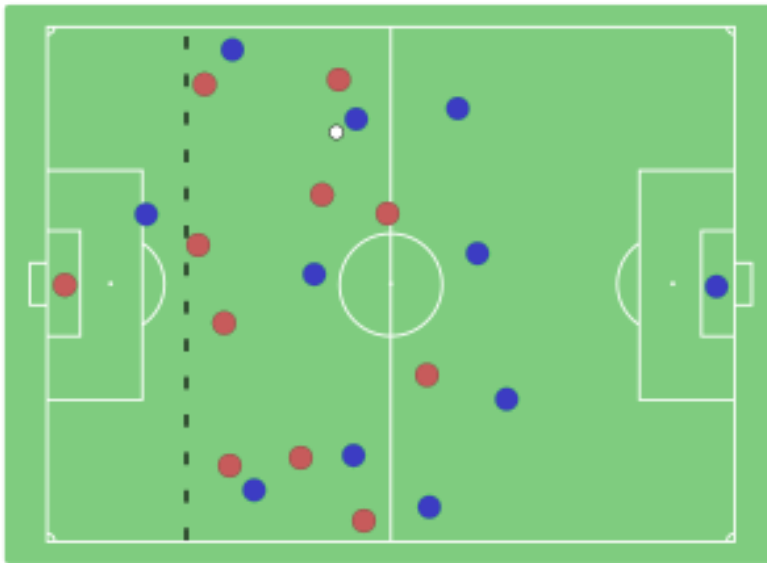
X, Y, Z 좌표

이미지 출처: FIFA

Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- 현실에 적용된 HAR 예시 (2022 카타르 월드컵)
 - 출력 값 : 반칙 조건에 충족 할 경우 오프사이드, 아닌 경우 온사이드

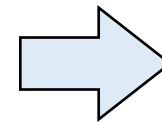
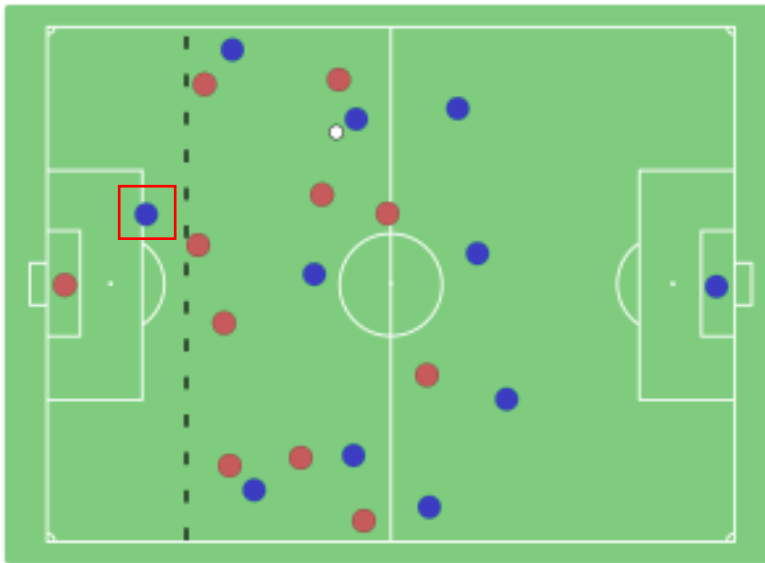


이미지 출처: 위키백과 오프사이드

Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- 현실에 적용된 HAR 예시 (2022 카타르 월드컵)
 - 출력 값 : 반칙 조건에 충족 할 경우 오프사이드, 아닌 경우 온사이드



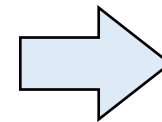
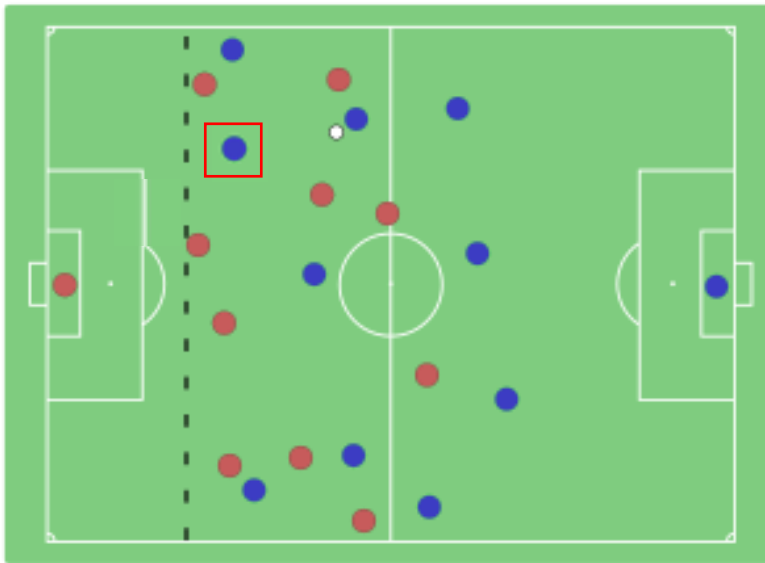
오프사이드

이미지 출처: 위키백과 오프사이드

Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- 현실에 적용된 HAR 예시 (2022 카타르 월드컵)
 - 출력 값 : 반칙 조건에 충족 할 경우 오프사이드, 아닌 경우 온사이드



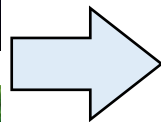
온사이드

이미지 출처: 위키백과 오프사이드

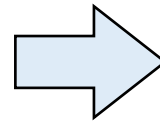
Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- 현실에 적용된 HAR 예시 (2022 카타르 월드컵)
 - 입력 값 1 : 12개의 카메라를 이용하여 사람의 위치를 추적 (50 FPS)
 - 입력 값 2 : 공에 들어있는 센서를 이용하여 사람의 위치를 추적 (500 FPS)
 - 출력 값 : 반칙 조건에 충족 할 경우 오프사이드, 아닌 경우 온사이드



학습모델

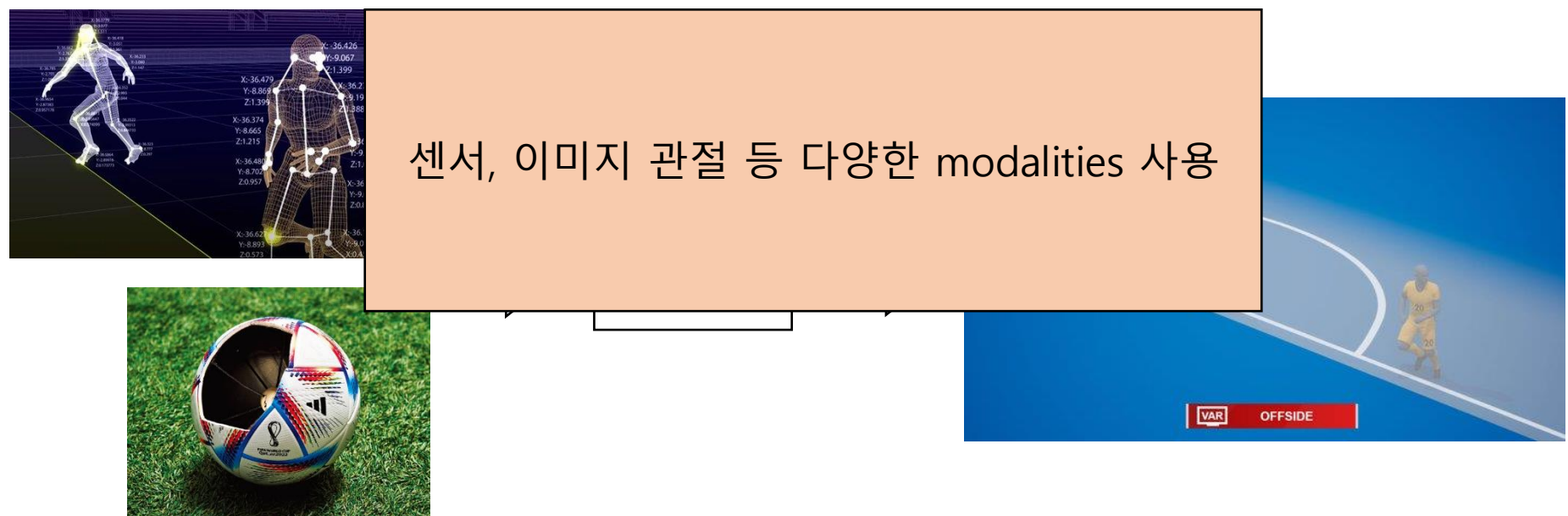


이미지 출처: FIFA

Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- 현실에 적용된 HAR 예시 (2022 카타르 월드컵)
 - 입력 값 1 : 12개의 카메라를 이용하여 사람의 위치를 추적 (50 FPS)
 - 입력 값 2 : 공에 들어있는 센서를 이용하여 사람의 위치를 추적 (500 FPS)
 - 출력 값 : 반칙 조건에 충족 할 경우 오프사이드, 아닌 경우 온사이드



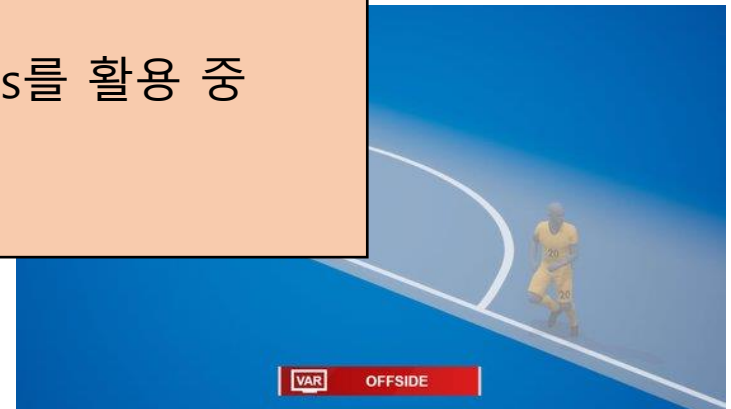
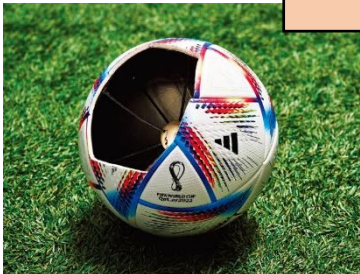
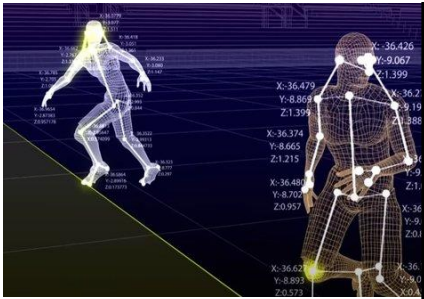
이미지 출처: FIFA

Introduction to Human Action Recognition

❖ Human Action Recognition (HAR)

- 현실에 적용된 HAR 예시 (2022 카타르 월드컵)
 - 입력 값 1 : 12개의 카메라를 이용하여 사람의 위치를 추적 (50 FPS)
 - 입력 값 2 : 공에 들어있는 센서를 이용하여 사람의 위치를 추적 (500 FPS)
 - 출력 값 : 반칙 조건에 충족 할 경우 오프사이드, 아닌 경우 온사이드

HAR에는 다양한 modalities를 활용 중

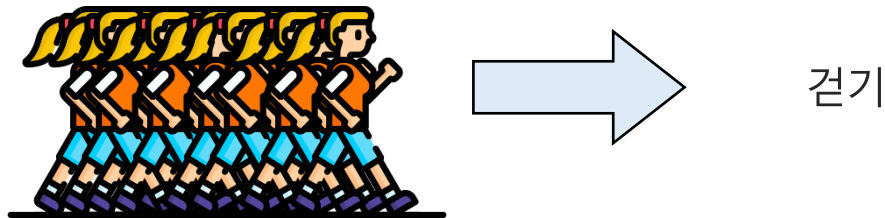


이미지 출처: FIFA

Human Action Recognition from single-modality

❖ HAR에서 목표로 하는 출력 값

- 일반적으로 연속적인 시점에서 하나의 행동을 추출하는 것을 목표로 함
 - 사용하는 시점의 개수는 데이터, 연구에 따라 다름
 - RGB, skeleton modality를 활용한 연구가 가장 활발하게 진행중



Human Action Recognition from single-modality

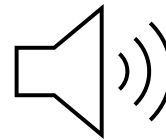
❖ HAR에서 사용되는 다양한 Modalities

- Visual modality와 non-visual modality로 분류 가능
 - Visual modality data : 카메라 기술을 기반으로 생성된 학습 데이터
 - Non-visual modality data : visual modality가 아닌 데이터



Visual modality data

VS



Non-Visual modality data

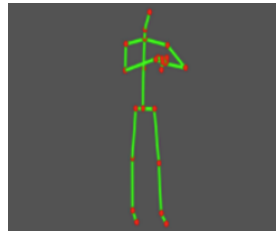
Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- Visual modality와 non-visual modality로 분류 가능
 - Visual modality data : 카메라 기술을 기반으로 생성된 학습 데이터
 - Non-visual modality data : visual modality가 아닌 데이터



RGB



Skeleton



Depth



Infrared



Point Cloud



Event Stream

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

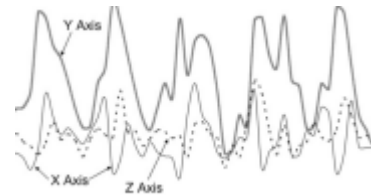
Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

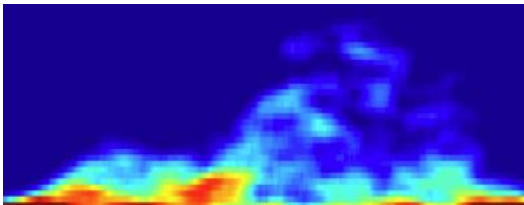
- Visual modality와 non-visual modality로 분류 가능
 - Visual modality data : 카메라 기술을 기반으로 생성된 학습 데이터
 - **Non-visual modality data : visual modality가 아닌 데이터**



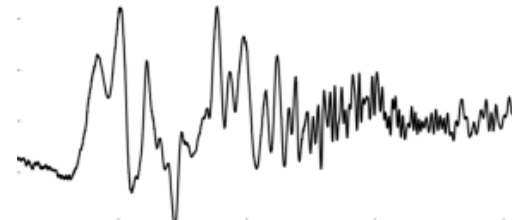
Audio



Acceleration



Radar



WiFi

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 각 Modality의 산출 방식과 장단점



장점

단점

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 각 Modality의 산출 방식과 장단점 (Visual modality data – RGB)
 - 일반적으로 수집되는 RGB로 구성된 이미지 데이터
 - 산출 방법 : RGB 카메라를 이용한 촬영



RGB

장점

풍부한 외형 정보 제공
상대적으로 데이터를 얻고 처리하는 과정이 간편
HAR이 아닌 다양한 분야에서도 활용 가능

단점

촬영 각도에 민감
배경에 민감
조명에 민감

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 각 Modality의 산출 방식과 장단점 (Visual modality data – Depth)
 - 카메라로부터 물체까지의 거리를 측정하여 생성한 흑백 이미지 데이터
 - 산출 방법 : 특수 카메라 사용



Depth

장점

구조 정보 제공
기하학적 형태의 정보 제공

단점

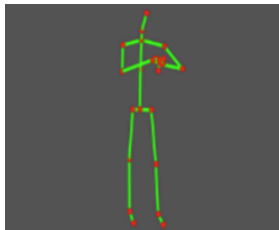
색상 및 텍스처 정보 부족
제한된 작업 가능 거리

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 각 Modality의 산출 방식과 장단점 (Visual modality data – Skeleton)
 - 사람의 관절 좌표와 뼈로 구성된 skeleton 데이터
 - 산출 방법 : RGB 이미지 혹은 depth 이미지에서 추론, 센서 사용, 특수 카메라 사용, 시뮬레이터 사용



Skeleton

장점

피사체 포즈의 구조 정보 제공
간단하면서도 유익한 정보
배경에 민감하지 않음

단점

외형 정보 부족
세부적인 형태 정보 부족
노이즈가 많은 편

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 각 Modality의 산출 방식과 장단점 (Visual modality data – Infrared sequence)
 - 적외선 카메라로 촬영된 이미지 데이터
 - 산출 방법 : 적외선 카메라 사용



Infrared

장점

어두운 환경에서도 작업 가능

단점

색상 및 질감 정보 부족
조명(특히 햇빛)에 취약함

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 각 Modality의 산출 방식과 장단점 (Visual modality data – Point Cloud)
 - 사람의 형태를 많은 점을 통해 표현한 데이터
 - 산출 방법 : 특수 카메라(LiDAR, Kinect) 사용, RGB 이미지 혹은 depth 이미지에서 추론



Point Cloud

장점

가장 구체적인 기하학적 모양으로 물체를 묘사
촬영 각도에 민감하지 않음

단점

색상 및 질감 정보 부족
높은 계산 복잡성

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 각 Modality의 산출 방식과 장단점 (Visual modality data – Event stream)
 - 배경과 같이 가만히 있는 부분보다 움직임이 있는 부분에 초점을 두어 촬영한 이미지
 - 산출 방법 : 특수 카메라 사용



Event Stream

장점

시각적으로 많은 중복성을 피할 수 있음

단점

시공간적 정보 부족
비싼 장치 가격

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 각 Modality의 산출 방식과 장단점 (Non-visual modality data – Audio)
 - 마이크를 통해 수집된 음성 데이터
 - 산출 방법 : 마이크 사용



Audio

장점

시점 내에서 작업을 쉽게 찾을 수 있음

단점

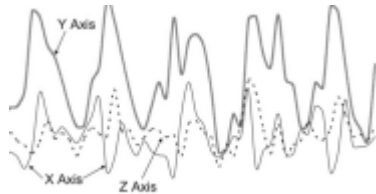
외형 정보 없음

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 각 Modality의 산출 방식과 장단점 (Non-visual modality data – Acceleration)
 - 가속도 센서를 통해 움직임을 숫자로 표현한 데이터
 - 산출 방법 : 피사체에 가속도계를 부착하여 측정



Acceleration

장점

세분화된 정보를 제공
개인 정보 보호
저렴한 비용

단점

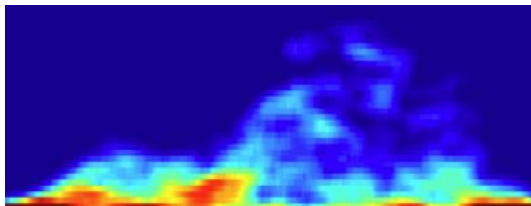
외형 정보 부족
피사체가 장치를 휴대해야 함

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 각 Modality의 산출 방식과 장단점 (Non-visual modality data – Radar)
 - 레이더 장치를 쏘아 얻어낸 반사파 데이터
 - 산출 방법 : 레이더 장치



Radar

장점

벽을 관통하여 사용 가능
조도에 민감하지 않음
날씨에 민감하지 않음
개인 정보 보호

단점

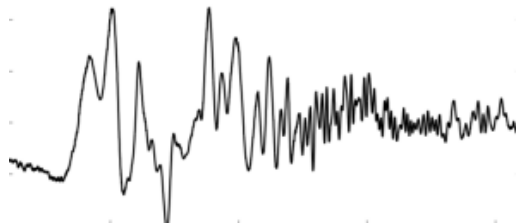
외형 정보 부족
비싼 장치 가격

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 각 Modality의 산출 방식과 장단점 (Non-visual modality data – WiFi)
 - 인체가 무선 신호를 반사한다는 점을 이용하여 수집
 - 산출 방법 : WiFi 신호



WiFi

장점

개인 정보 보호
저렴한 비용

단점

외관 정보 부족
환경에 민감
노이즈 존재

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- Visual modality와 non-visual modality로 분류 가능
 - Visual modality data : 카메라 기술을 기반으로 생성된 학습 데이터
 - Non-visual modality data : visual modality가 아닌 데이터



RGB



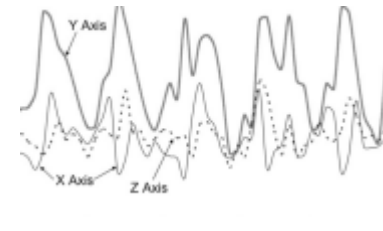
Skeleton



Depth



Audio



Acceleration



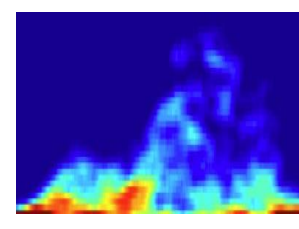
Infrared



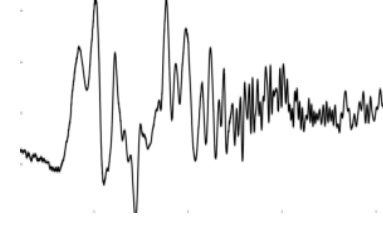
Point Cloud



Event Stream



Radar



WiFi

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- Visual modality와 non-visual modality로 분류 가능
 - Visual modality data : 카메라 기술을 기반으로 생성된 학습 데이터
 - Non-visual modality data : visual modality가 아닌 데이터



RGB



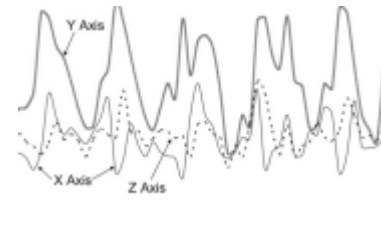
Skeleton



Depth



Audio



Acceleration



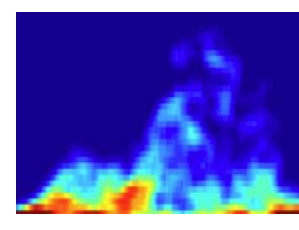
Infrared



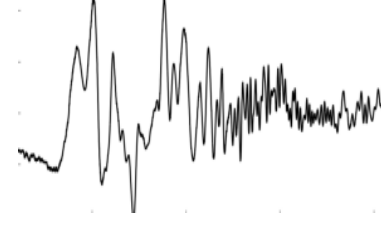
Point Cloud



Event Stream



Radar



WiFi

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

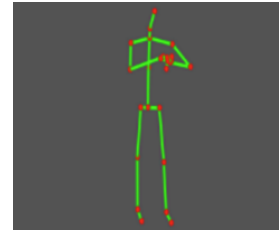
❖ HAR에서 사용되는 다양한 Modalities

- Visual modality와 non-visual modality로 분류 가능
 - Visual modality data : 카메라 기술을 기반으로 생성된 학습 데이터
 - Non-visual modality data : visual modality가 아닌 데이터

Convolution



RGB



Skeleton

Graph

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

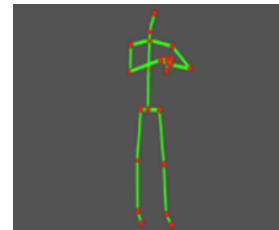
❖ HAR에서 사용되는 다양한 Modalities

- Visual modality와 non-visual modality로 분류 가능
 - Visual modality data : 카메라 기술을 기반으로 생성된 학습 데이터
 - Non-visual modality data : visual modality가 아닌 데이터

3D CNN for HAR



RGB



Skeleton

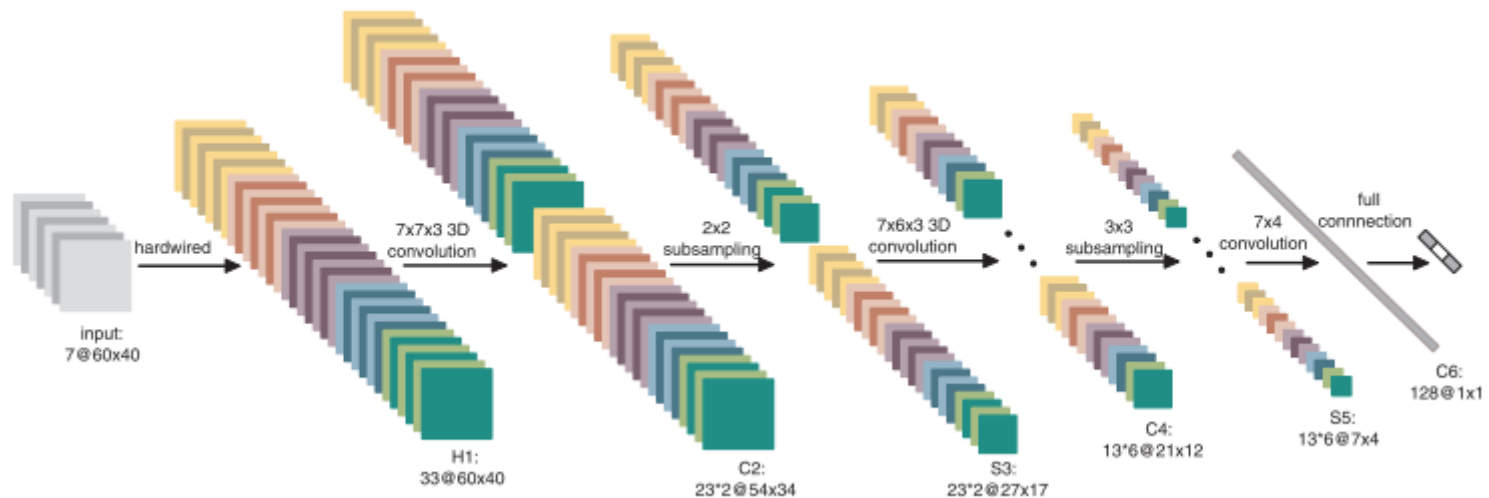
ST-GCN for skeleton-based action recognition

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 3D Convolutional Neural Networks for Human Action Recognition (RGB)
 - IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, citations: 6189

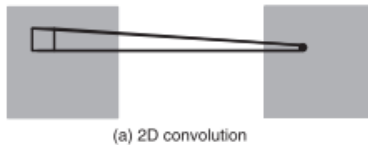


Ji, S., Xu, W., Yang, M., & Yu, K. (2012). 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(1), 221-231.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 3D Convolutional Neural Networks for Human Action Recognition (RGB)
 - IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, citations: 6189

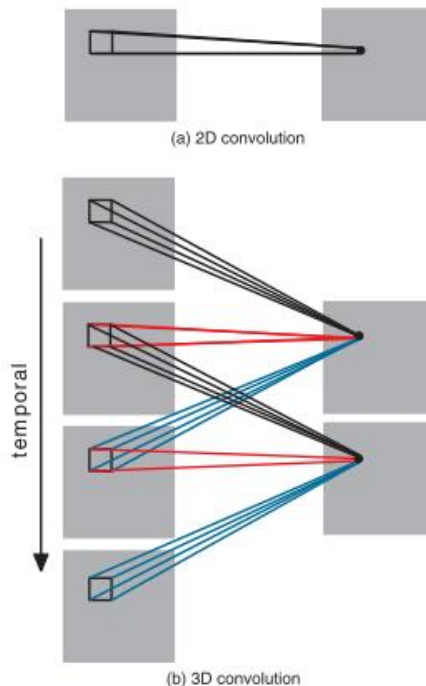


Ji, S., Xu, W., Yang, M., & Yu, K. (2012). 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(1), 221-231.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 3D Convolutional Neural Networks for Human Action Recognition (RGB)
 - IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, citations: 6189

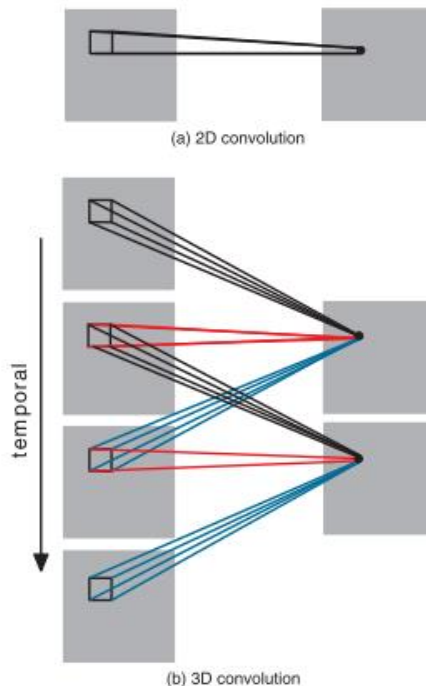


Ji, S., Xu, W., Yang, M., & Yu, K. (2012). 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(1), 221-231.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- 3D Convolutional Neural Networks for Human Action Recognition (RGB)
 - IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, citations: 6189



2D inputs : channel, width, height

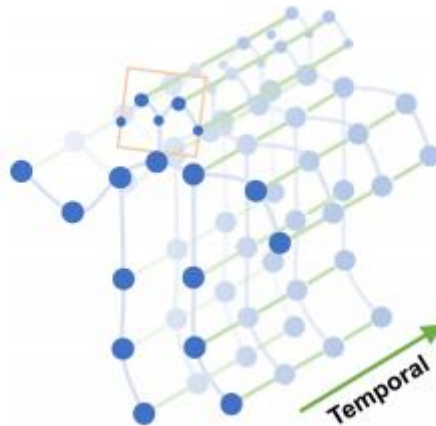
3D inputs : **temporal**, channel, width, height

Ji, S., Xu, W., Yang, M., & Yu, K. (2012). 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(1), 221-231.

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- Spatial Temporal Graph Convolutional Neural Networks for Skeleton-Based Action Recognition (Skeleton)
 - AAAI Conference on Artificial Intelligence, 2018, citations: 2751

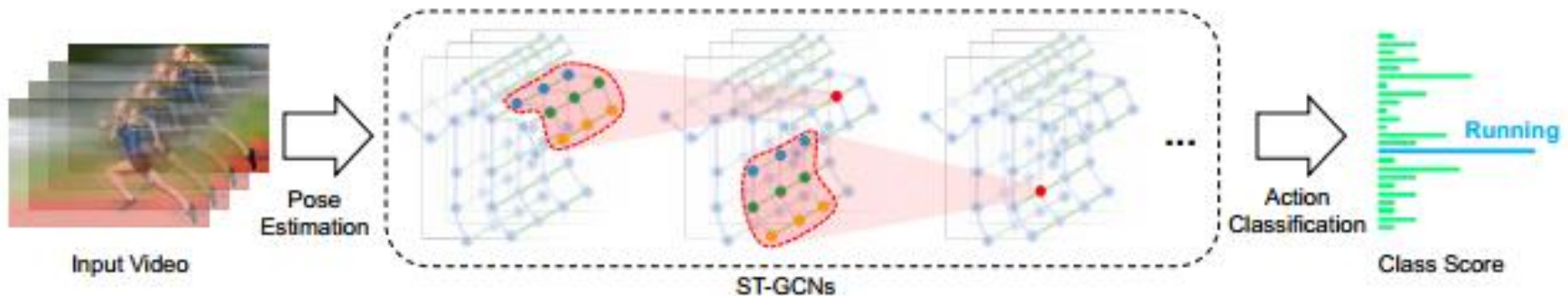


Yan, S., Xiong, Y., & Lin, D. (2018, April). Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 32, No. 1).

Human Action Recognition from single-modality

❖ HAR에서 사용되는 다양한 Modalities

- Spatial Temporal Graph Convolutional Neural Networks for Skeleton-Based Action Recognition (Skeleton)
 - AAAI Conference on Artificial Intelligence, 2018, citations: 2751



Yan, S., Xiong, Y., & Lin, D. (2018, April). Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 32, No. 1).

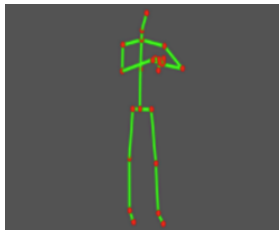
Human Action Recognition from multi-modalities

❖ HAR에서 사용되는 다양한 Modalities

- Multi-modalities 결합 학습 방식
 - 다양한 modalities가 인공지능 학습에 잘 활용되고 있다는 것을 확인
 - 반면, 각각의 modality는 다른 특징을 가지고 있음
 - 이러한 점을 생각하여 두가지 이상을 활용한 학습을 진행하고자 함



RGB



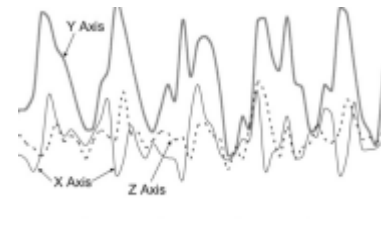
Skeleton



Depth



Audio



Acceleration



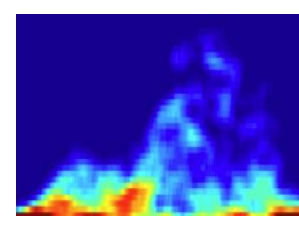
Infrared



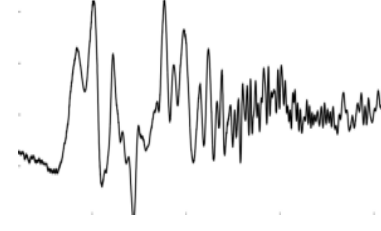
Point Cloud



Event Stream



Radar



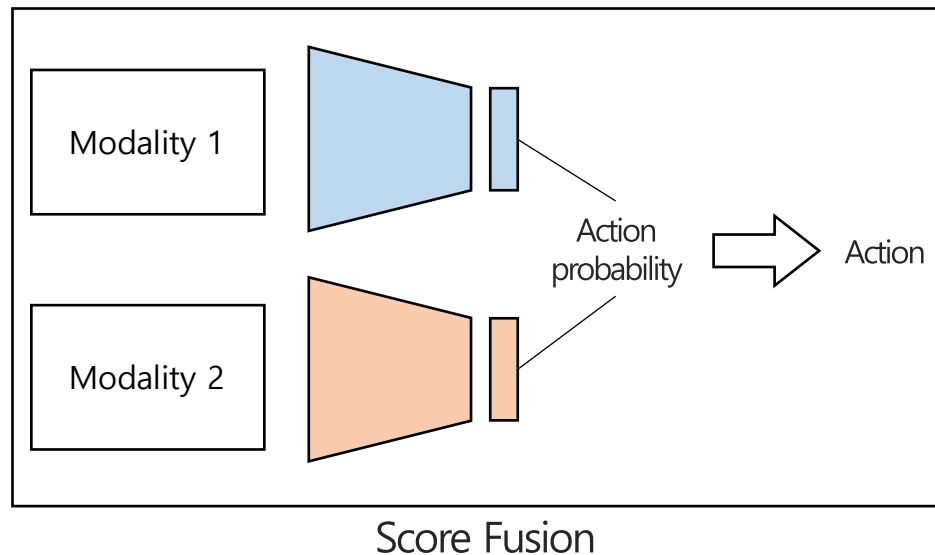
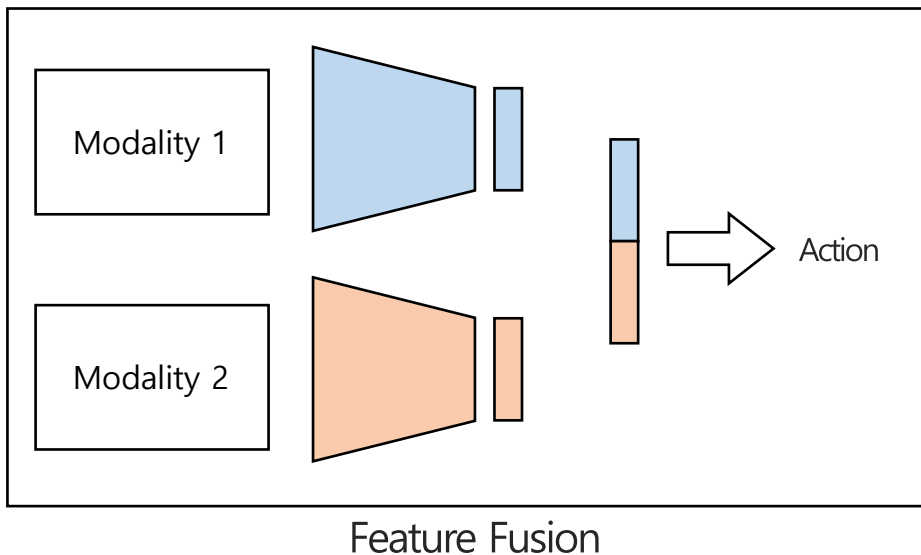
WiFi

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from multi-modalities

❖ HAR에서 사용되는 다양한 Modalities

- Multi-modalities 결합 학습 방식
 - Feature : 두개 이상의 Modalities의 특징들을 모델 연산 중 결합하는 방식
 - Score : 두개 이상의 Modalities를 각각 학습 후 결합하는 방식



Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Human Action Recognition from multi-modalities

❖ HAR에서 사용되는 다양한 Modalities

- Human Action Recognition without Human

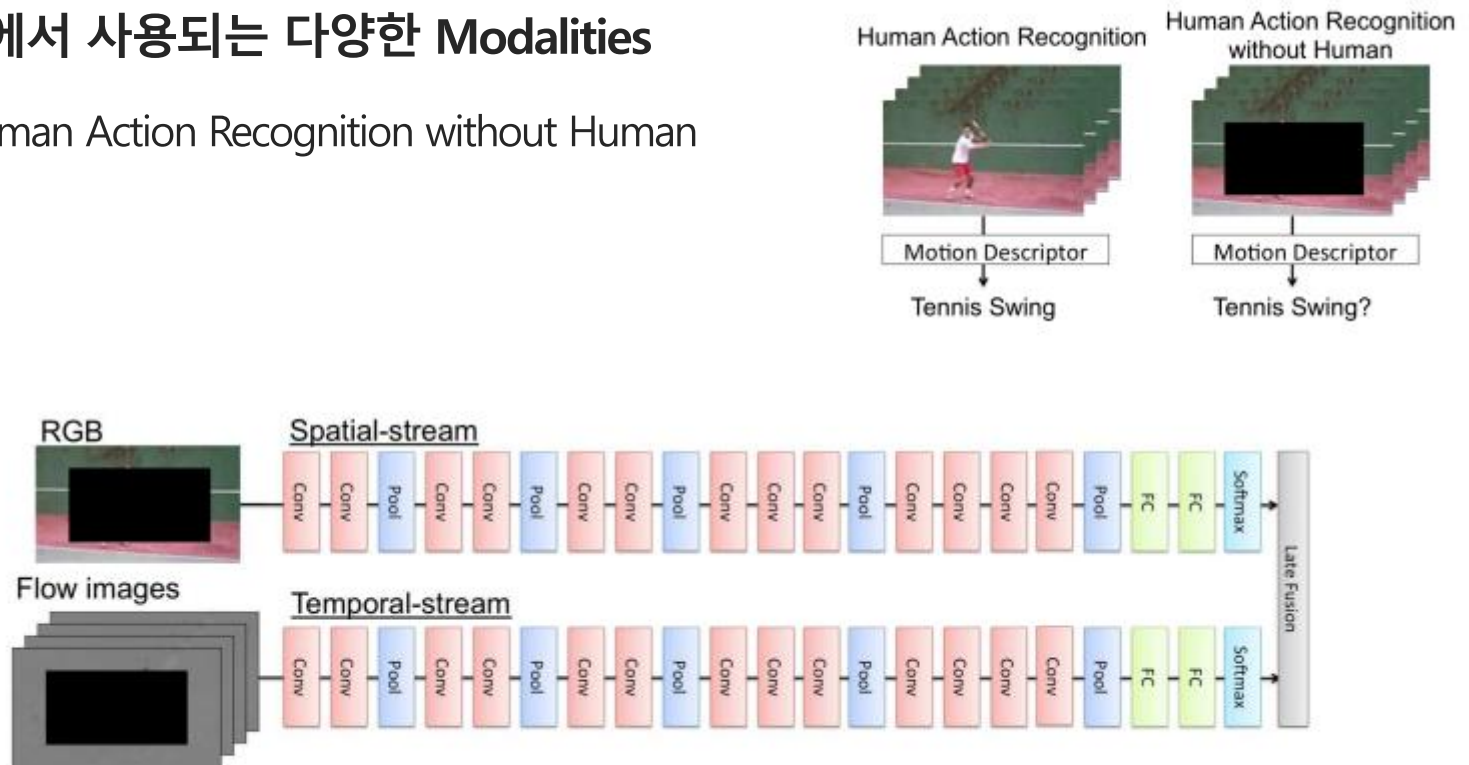


Figure 2: Very deep two-stream CNN [19] for human action recognition without human.

He, Y., Shirakabe, S., Satoh, Y., & Kataoka, H. (2016). Human action recognition without human. In *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14* (pp. 11-17). Springer International Publishing.

Human Action Recognition from multi-modalities

❖ HAR에서 사용되는 다양한 Modalities

- Human Action Recognition without Human

With or Without a Human	Stream	% on UCF101 (split 1)
With human	Spatial stream	51.26
	Temporal stream	40.50
	Two-stream	56.91
Without human	Spatial stream	45.33
	Temporal stream	26.80
	Two-stream	47.42

Table 2: Performance rate of human action recognition with or without a human

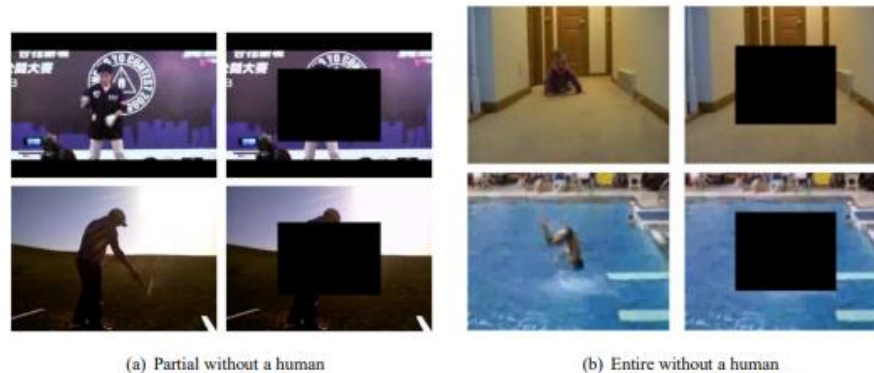


Figure 4: Qualitative evaluation of the setting without a human on the UCF101 dataset.

He, Y., Shirakabe, S., Satoh, Y., & Kataoka, H. (2016). Human action recognition without human. In *Computer Vision—ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14* (pp. 11-17). Springer International Publishing.

Conclusion

❖ Human Action Recognition (HAR)

- HAR은 다양한 modality 입력 값을 사용하여 사람의 현재 행동을 인식하는 분야
 - Visual vs Non-visual modality
- Single-modality를 활용한 연구의 경우 다양한 딥러닝 기반 연구가 다수 존재
 - RNN, CNN, GCN, Transformer (RGB 및 skeleton이 가장 많이 활용)
- 다양한 modality가 존재하는 만큼 multi-modalities를 활용한 학습 연구가 많이 진행중
 - Feature 또는 score 단계에서 모델을 합성하여 학습을 수행
- 현실에서 활용 가능성이 높은 분야
 - 실제로 범죄 예방, 스포츠, 게임 등에서 활용 중

감사합니다