
Skeleton-based Human Activity Recognition

2022. 02. 04.

고려대학교

Data Mining and Quality Analytics

발표자 소개

❖ 조용원(E-mail: gyj4318@korea.ac.kr)

- 현재 석·박사 통합과정 6학기 재학 중 (지도교수: 김성범)
- 연구분야
 - ✓ Supervised Semantic Segmentation
 - ✓ Visual Anomaly Detection and Segmentation
 - ✓ Skeleton-based Human Action Recognition



참고 내용 1

종료

인간의 눈을 대체할 수 있는 **시각 기능**(사람이 어디에 있는지에 대한 위치 정보)

- 시각 기능을 통해 획득한 **자세에 대한 정보**(관절 중심점에 대한 위치정보)
- 올바른 베드민턴 자세에 대한 지식

Introduction to Human Pose Estimation

발표자: 조용원

📅 2021년 2월 26일
🕒 오후 1시 ~
📺 온라인 비디오 시청 (YouTube)

세미나 정보 보기 →

참고 내용 2

종료

Intermediate Human Pose Estimation

2021. 08. 27

Data Mining and Quality Analytics Lab

기사미

Intermediate Human Pose Estimation

발표자: 김상민

📅 2021년 8월 27일
🕒 오후 12시 ~
📺 온라인 비디오 시청 (YouTube)

세미나 정보 보기 →

목차

- ❖ Introduction
- ❖ Spatial-temporal Graph Convolutional Networks
- ❖ Applications for Skeleton-based HAR
- ❖ Conclusion

Introduction

- Motivation

❖ 갑자기 인간이 쓰러졌을 때, 자동적으로 인식할 수 있을까?

- 위급한 상황 발생 시, 이를 자동으로 인식할 수 있을까?
- 폐쇄형 텔레비전(Closed-circuit television, CCTV)에서 이상 상황을 자동 감지 가능할 까?
- 많은 CCTV를 소수 인원이 감시하는 경우, 제대로 CCTV 이상 상황 인식이 가능할 까?

엘리베이터 내에서 사람이 쓰러진 경우



CCTV 관제센터



- <https://www.seoul.co.kr/news/newsView.php?id=20170327010015>
- <http://www.cctvnews.co.kr/news/articleView.html?idxno=54520>

Introduction

- Human activity recognition

❖ 인간 행동 인식(Human activity recognition, HAR)이란?

- 스마트 시계 사용 활성화로 인해 인간 활동에 대한 **시그널 정보** 수집
 - 스마트 시계 내 동작 센서, 산소 포화도 센서, 피부 전도 센서 등 존재
- 다양한 분야에서 CCTV 를 사용해 **영상 정보** 수집
 - 공공(치안), 제조(작업장), 유통(대형마트) 등 다양한 분야에서 CCTV 영상 수집
- **시그널** 또는 **영상**을 입력 받아 **인간이 어떠한 행동을 하고 있는 지**를 인식하는 문제

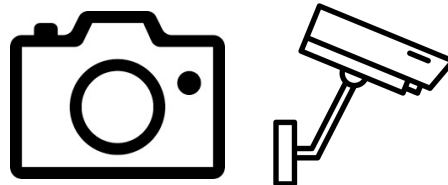
인간 행동 상황



시그널 정보 수집



영상 정보 수집



인공지능
알고리즘

타격
송구
수비
.....

- <https://www.itworld.co.kr/news/88157>

- <https://m.post.naver.com/viewer/post/view.nhn?volumeNo=8637181&memberNo=36733075>

Introduction

- Human activity recognition

❖ 인간 행동 인식(Human activity recognition, HAR)이란?

- 스마트 시계 사용 활성화로 인해 인간 활동에 대한 **시그널 정보** 수집
 - 스마트 시계 내 동작 센서, 산소 포화도 센서, 피부 전도 센서 등 존재
- 다양한 분야에서 CCTV 를 사용해 **영상 정보** 수집
 - 공공(치안), 제조(작업장), 유통(대형마트) 등 다양한 분야에서 CCTV 영상 수집
- **시그널** 또는 **영상**을 입력 받아 **인간이 어떠한 행동을 하고 있는 지**를 인식하는 문제



- <https://www.itworld.co.kr/news/88157>

- <https://m.post.naver.com/viewer/post/view.nhn?volumeNo=8637181&memberNo=36733075>

Introduction

- Taxonomy of HAR

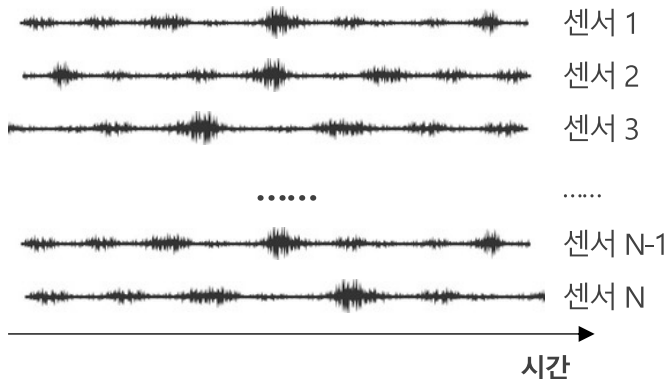
❖ 인간 행동 인식(HAR) 계층도

- 입력 데이터 형태에 따른 분류 가능
- 영상 데이터 입력 방식에 따른 분류



Human Activity Recognition (HAR)

Sensor-based HAR



행동 유형 인식

Vision-based HAR

영상



Introduction

- Taxonomy of HAR

❖ 인간 행동 인식(HAR) 계층도

- 입력 데이터 형태에 따른 분류 가능
- 영상 데이터 입력 방식에 따른 분류



타격
송구
수비
.....

Human Activity Recognition (HAR)

Sensor-based HAR

Vision-based HAR

영상

Video-based HAR

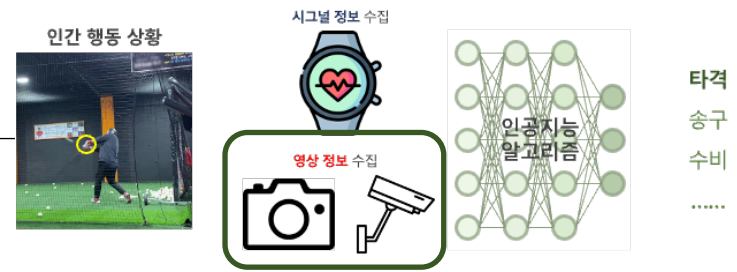


Introduction

- Taxonomy of HAR

❖ 인간 행동 인식(HAR) 계층도

- 입력 데이터 형태에 따른 분류 가능
- 영상 데이터 입력 방식에 따른 분류



Human Activity Recognition (HAR)

Sensor-based HAR

Vision-based HAR

영상 내 프레임



Human Pose Estimation 결과



Video-based HAR

Skeleton-based HAR

Introduction

- Taxonomy of HAR



❖ 인간 행동 인식(HAR) 계층도

- 입력 데이터 형태에 따른 분류 가능
- 영상 데이터 입력 방식에 따른 분류

Human Activity Recognition (HAR)

Sensor-based HAR

Vision-based HAR

Video-based HAR

Skeleton-based HAR

첫번째 프레임에
대한 Skeleton

두번째 프레임에
대한 Skeleton

세번째 프레임에
대한 Skeleton

네번째 프레임에
대한 Skeleton



Introduction

- Taxonomy of HAR



❖ 인간 행동 인식(HAR) 계층도

- 입력 데이터 형태에 따른 분류 가능
- 영상 데이터 입력 방식에 따른 분류

Human Activity Recognition (HAR)

Sensor-based HAR

Vision-based HAR

Video-based HAR

Skeleton-based HAR

첫번째 프레임에
대한 Skeleton

두번째 프레임에
대한 Skeleton

세번째 프레임에
대한 Skeleton

네번째 프레임에
대한 Skeleton



Introduction

- Importance of Skeleton-based HAR

❖ 왜 Skeleton-based HAR 방식으로 접근해야 하는가?

- Video-based HAR은 다양한 영상 데이터를 수집해야 한다는 단점
 - 촬영 각도, 배경과 체구 변화에 따라 Video-based HAR 성능 변화가 심하다고 알려짐
- Skeleton-based HAR은 여러 변화에도 강건한 성능을 보유
 - Skeleton은 촬영 각도, 배경 관련 요소가 존재하지 않음
 - 자세 자체(Skeleton)를 입력 데이터로 사용하기 때문에 행동에 대한 좋은 특징 추출 가능

166cm 사람 스윙



172cm 사람 스윙



166cm 사람 스윙(배경 변화)



Introduction

- Importance of Skeleton-based HAR

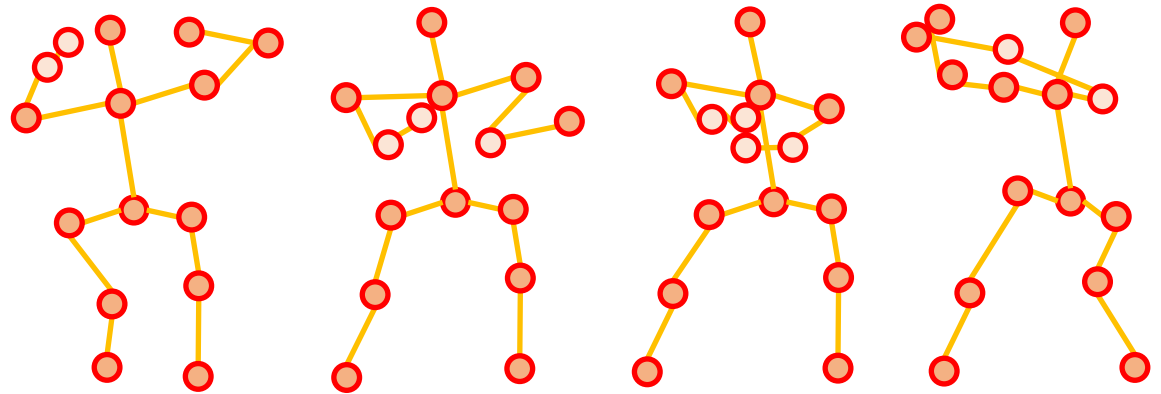
❖ 왜 Skeleton-based HAR 방식으로 접근해야 하는가?

- Video-based HAR은 다양한 영상 데이터를 수집해야 한다는 단점
 - 촬영 각도, 배경과 체구 변화에 따라 Video-based HAR 성능 변화가 심하다고 알려짐
- Skeleton-based HAR은 여러 변화에도 강건한 성능을 보유
 - Skeleton은 **촬영 각도, 배경 관련 요소**가 존재하지 않음
 - 자세 자체(Skeleton)를 입력 데이터로 사용하기 때문에 행동에 대한 **좋은 특징 추출 가능**

166cm 사람 스윙



166cm 사람 스윙에 대한 Skeleton



목차

- ❖ Introduction
- ❖ **Spatial-temporal Graph Convolutional Networks**
- ❖ Applications for Skeleton-based HAR
- ❖ Conclusion

Spatial Temporal Graph Convolutional Networks

❖ Spatial Temporal Graph Convolutional Networks for Skeleton-based AR

- 2018년 Association for the Advancement of Artificial Intelligence (AAAI)에서 발표
- 저자들은 홍콩 중문대학 소속이며 22년 2월 3일 기준 1,650회 인용된 논문
- Skeleton을 Graph로 정의하고 Graph convolutional networks(GCN) 기반 분류 모델 제안

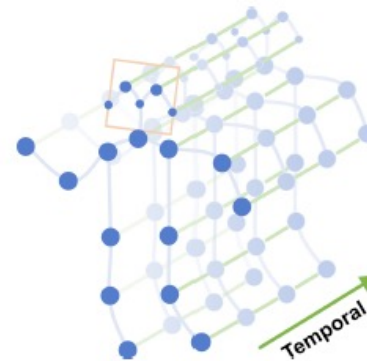
Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition

Sijie Yan, Yuanjun Xiong, Dahua Lin

Department of Information Engineering, The Chinese University of Hong Kong
{ys016, dhlin}@ie.cuhk.edu.hk, bitxiong@gmail.com

Abstract

Dynamics of human body skeletons convey significant information for human action recognition. Conventional approaches for modeling skeletons usually rely on hand-crafted parts or traversal rules, thus resulting in limited expressive power and difficulties of generalization. In this work, we propose a novel model of dynamic skeletons called *Spatial-Temporal Graph Convolutional Networks (ST-GCN)*, which moves beyond the limitations of previous methods by automatically learning both the spatial and temporal patterns from data. This formulation not only leads to greater expressive power but also stronger generalization capability. On two large datasets, *Kinetics* and *NTU-RGBD*, it achieves substantial improvements over mainstream methods.

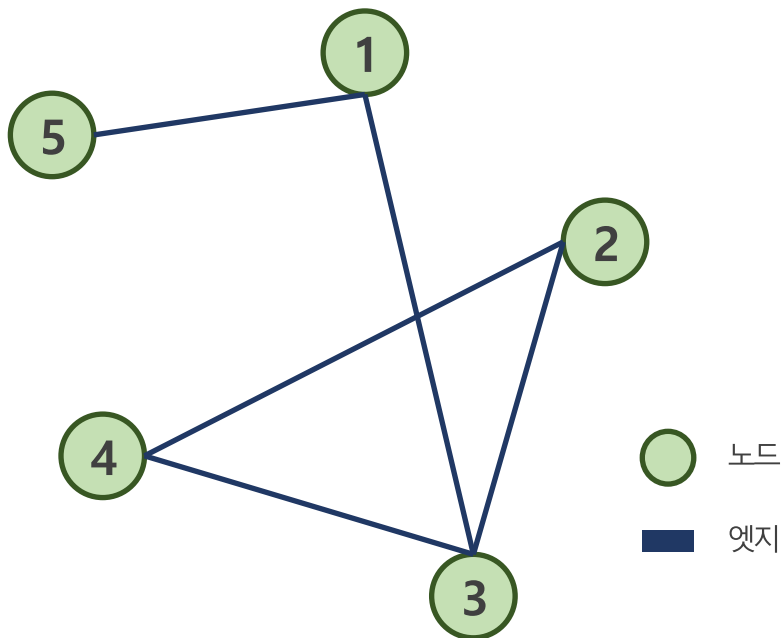


Spatial Temporal Graph Convolutional Networks

- Definition of Graph

❖ Graph 데이터 정의

- Graph 데이터는 여러 점(Vertex, Node)과 선(Edge)의 집합
- 데이터 간의 관계를 표현할 때 사용하는 데이터 형태
- 화학 분자 구조, 인용 관계, 3D Mesh 데이터 등 다양한 분야에서 Graph 데이터 사용

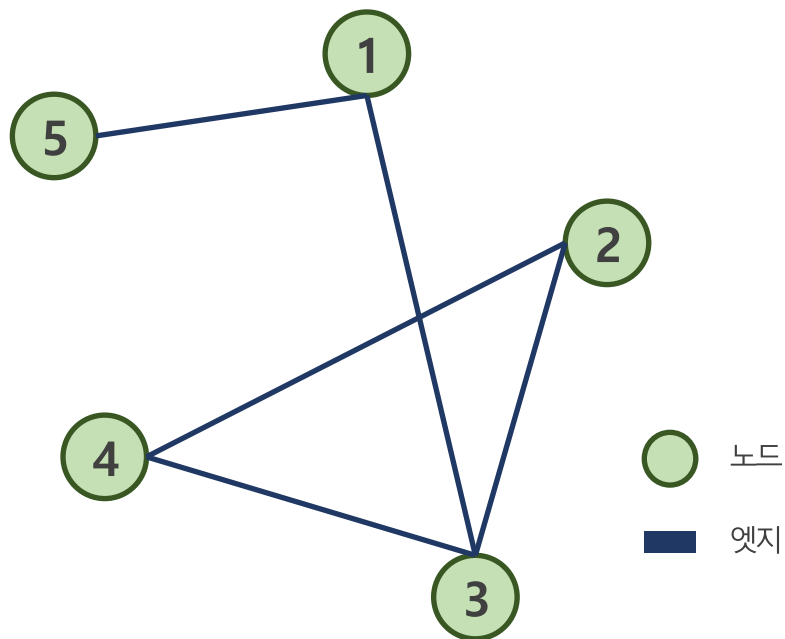


Spatial Temporal Graph Convolutional Networks

- Definition of Graph

❖ Graph 데이터 정의

- Graph 데이터는 여러 점(Vertex, Node)과 선(Edge)의 집합
- 데이터 간의 관계를 표현할 때 사용하는 데이터 형태
- 화학 분자 구조, 인용 관계, 3D Mesh 데이터 등 다양한 분야에서 Graph 데이터 사용



노드 특징 벡터
(Node feature vector)

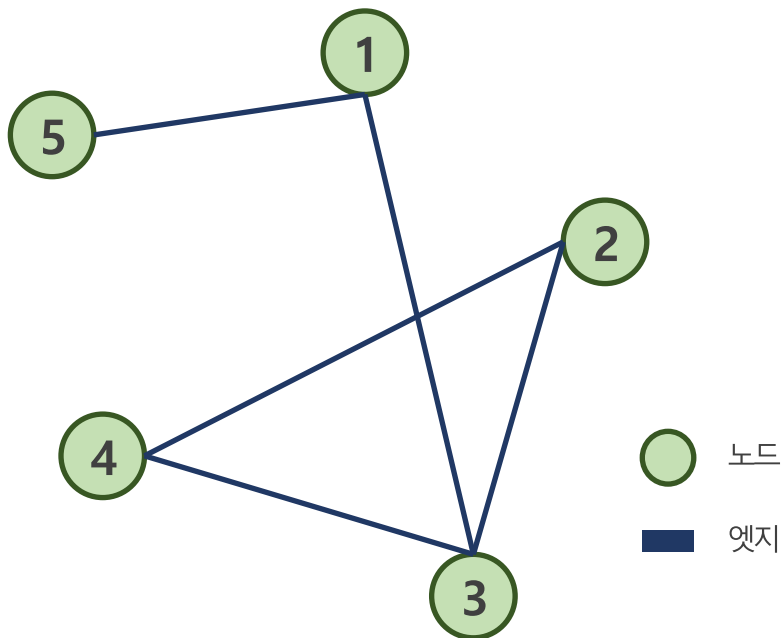
1	변수 1	변수 2	변수 3	변수 4
2	변수 1	변수 2	변수 3	변수 4
.....				
5	변수 1	변수 2	변수 3	변수 4

Spatial Temporal Graph Convolutional Networks

- Definition of Graph

❖ Graph 데이터 정의

- Graph 데이터는 여러 점(Vertex, Node)과 선(Edge)의 집합
- 데이터 간의 관계를 표현할 때 사용하는 데이터 형태
- 화학 분자 구조, 인용 관계, 3D Mesh 데이터 등 다양한 분야에서 Graph 데이터 사용



인접 행렬
(Adjacency matrix)

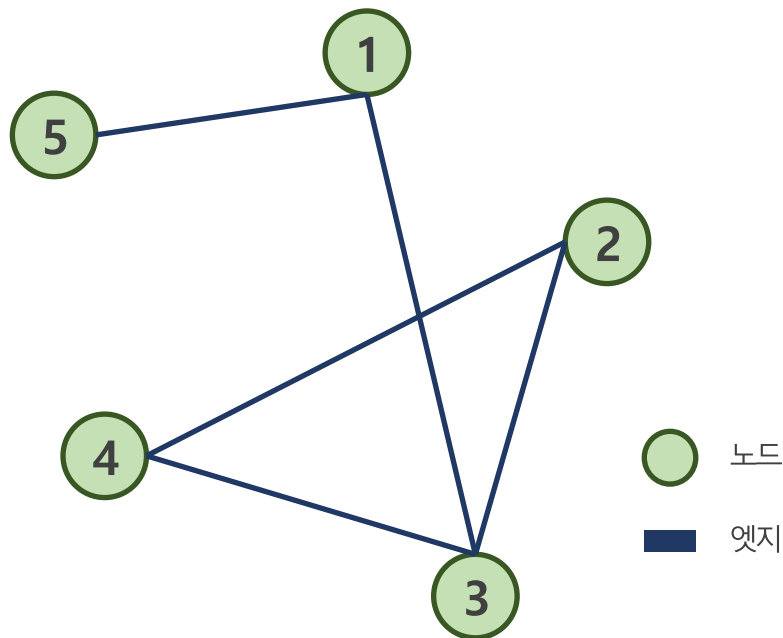
	1	2	3	4	5
1			1		1
2					
3	1				
4					
5	1				

Spatial Temporal Graph Convolutional Networks

- Definition of Graph

❖ Graph 데이터 정의

- Graph 데이터는 여러 점(Vertex, Node)과 선(Edge)의 집합
- 데이터 간의 관계를 표현할 때 사용하는 데이터 형태
- 화학 분자 구조, 인용 관계, 3D Mesh 데이터 등 다양한 분야에서 Graph 데이터 사용



인접 행렬
(Adjacency matrix)

	1	2	3	4	5
1	0	0	1	0	1
2	0	0	1	1	0
3	1	1	0	1	0
4	0	1	1	0	0
5	1	0	0	0	0

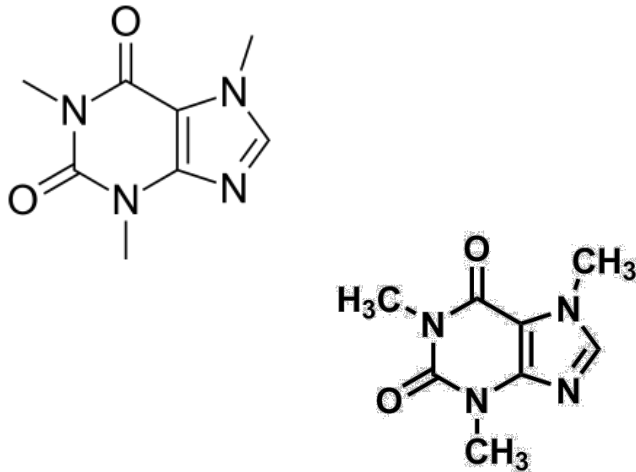
Spatial Temporal Graph Convolutional Networks

- Definition of Graph

❖ Graph 데이터 정의

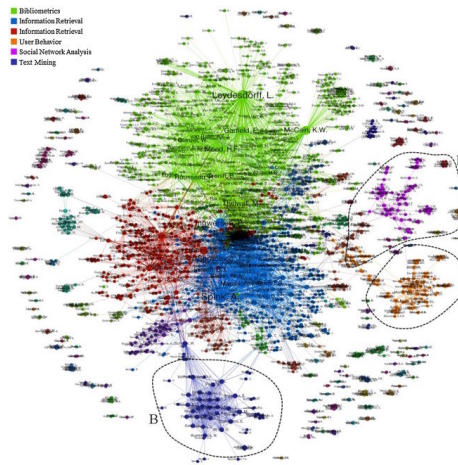
- Graph 데이터는 여러 점(Vertex, Node)과 선(Edge)의 집합
- 데이터 간의 관계를 표현할 때 사용하는 데이터 형태
- 화학 분자 구조, 인용 관계, 3D Mesh 데이터 등 다양한 분야에서 Graph 데이터 사용

화학 분자 구조
(Molecular graph)



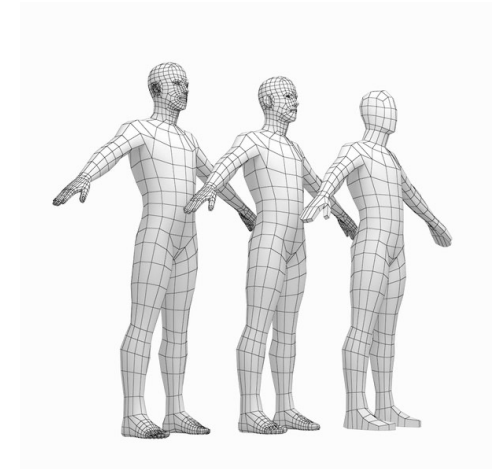
노드: 화학 원소 / 엣지: 원소 간 연결 정보

인용 관계
(Citation graph)



노드: 논문 명 / 엣지: 인용 관계

3D Mesh 형태
(일종의 Polygon)



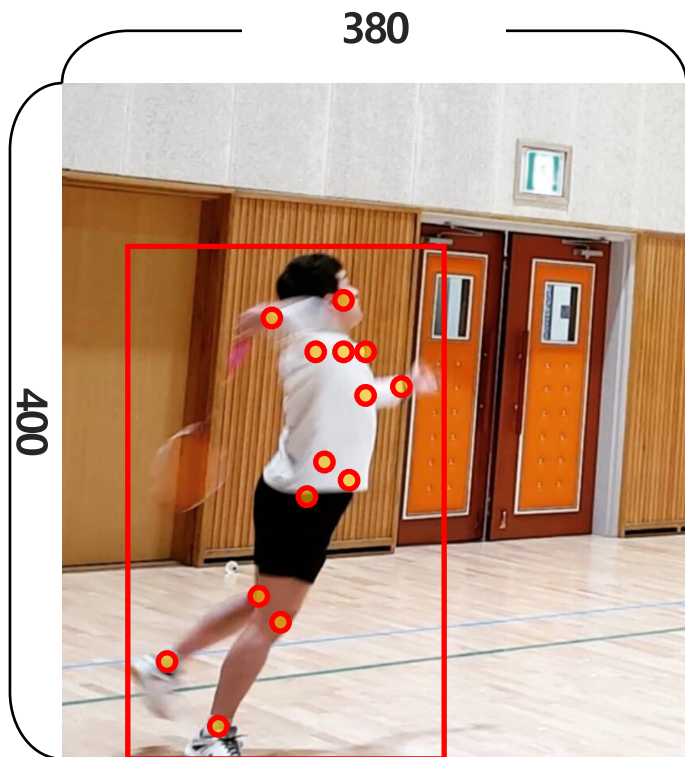
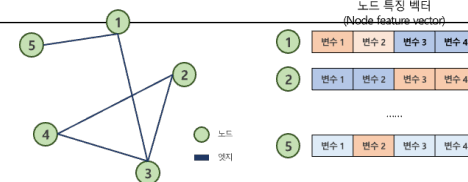
노드: 꼭지점 / 엣지: 점 사이 연결

Spatial Temporal Graph Convolutional Networks

- Human Pose Estimation

❖ 인간 자세 추정(Human Pose estimation, HPE)

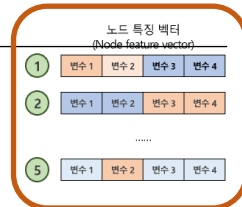
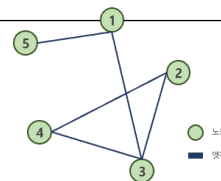
- 사진이나 영상을 입력 받아 **인간의 주요 관절에 대한 좌표 값**을 예측하는 문제
- 관절(노드)별 좌표 값 뿐만 아니라 사진/영상 내 관절이 보이는 지 여부에 대한 **확률**도 추론
- Skeleton-based HAR에서는 OpenPose 모델의 예측 결과를 입력 데이터로 사용



- ① 머리: (120, 370, Z축 좌표, 1.00)
- ② 왼쪽 어깨: (120, 330, Z축 좌표, 0.63)
- ③ 왼쪽 팔꿈치: (100, 330, Z축 좌표, 0.53)
- ④ 왼쪽 손목: (80, 350, Z축 좌표, 0.80)
- ⋮
- ⑫ 오른쪽 발목: (50, 70, Z축 좌표, 1.00)

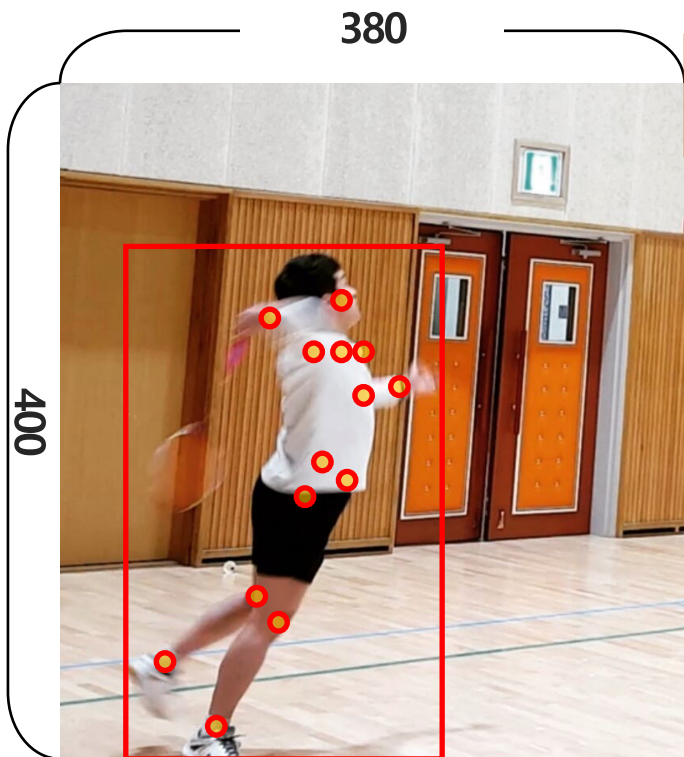
Spatial Temporal Graph Convolutional Networks

- From Skeleton to Graph



❖ 인간 자세 추정(Human Pose estimation, HPE)

- 사진이나 영상을 입력 받아 **인간의 주요 관절에 대한 좌표 값**을 예측하는 문제
- 관절(노드)별 좌표 값 뿐만 아니라 사진/영상 내 관절이 보이는 지 여부에 대한 **확률**도 추론
- Skeleton-based HAR에서는 OpenPose 모델의 예측 결과를 입력 데이터로 사용



2D Skeleton-based HAR: 관절 별 (X 좌표, Y 좌표, **확률**)
3D Skeleton-based HAR: 관절 별 (X 좌표, Y 좌표, Z 좌표, **확률**)

- ① 머리: (120, 370, Z축 좌표, **1.00**)
- ② 왼쪽 어깨: (120, 330, Z축 좌표, **0.63**)
- ③ 왼쪽 팔꿈치: (100, 330, Z축 좌표, **0.53**)
- ④ 왼쪽 손목: (80, 350, Z축 좌표, **0.80**)
- ⋮
- ⑫ 오른쪽 발목: (50, 70, Z축 좌표, **1.00**)

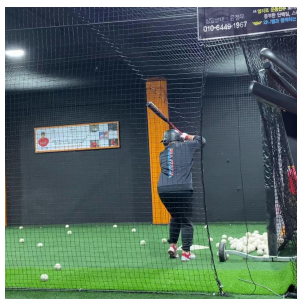
Spatial Temporal Graph Convolutional Networks

- From Skeleton to Graph

❖ 프레임 별 Skeleton 을 연결해 단일 Graph로 변환

- 영상 내 프레임 별 Skeleton을 산출(Intra-body connection)
 - 인간 관절 사이에 어떠한 관절 사이 연결되어야 하는지는 정해져 있음(Intra-body connection 정의)
- 프레임 사이 동일한 관절에 대한 연결 생성(Inter-frame connection)

프레임 1



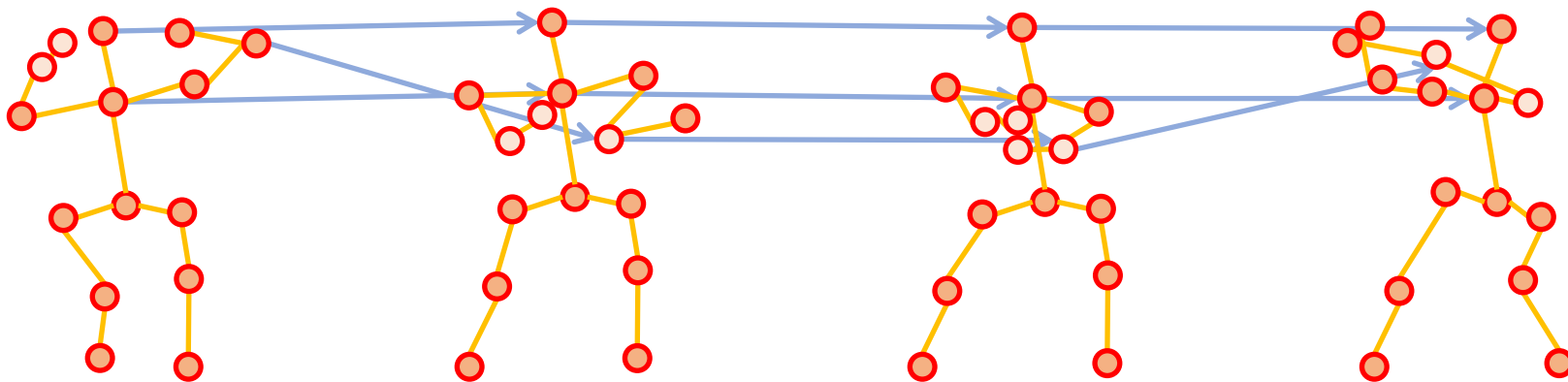
프레임 2



프레임 3



프레임 4



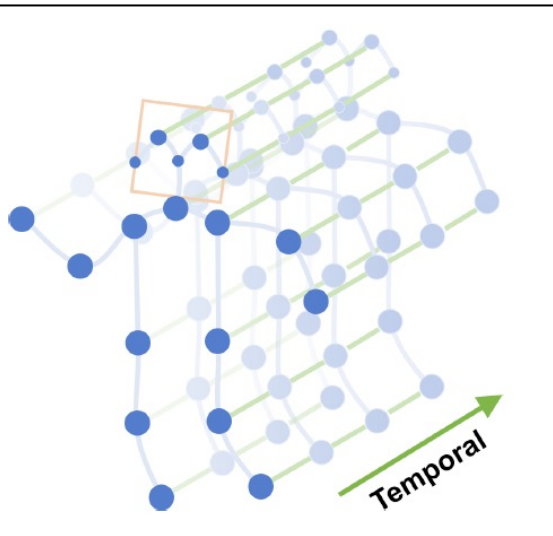
Spatial Temporal Graph Convolutional Networks

- ST-GCN

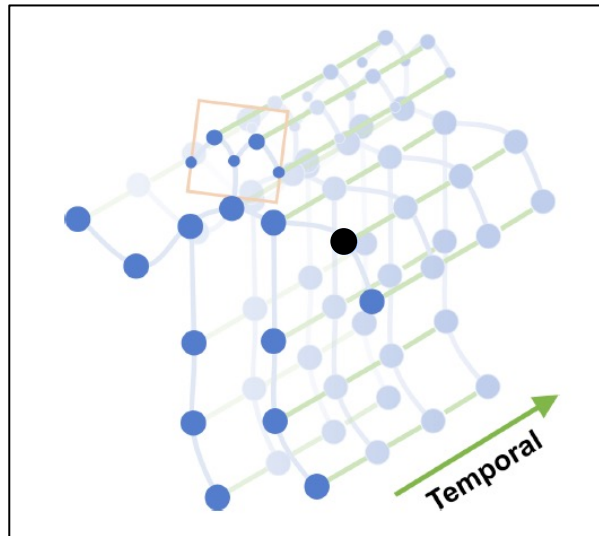
❖ Spatial Temporal Graph Convolutional Networks(ST-GCN)

- 개별 영상 내 모든 프레임에 대해 연결 진행(Intra-body and Inter-frame connection)
- 그래프 내 **특정 관절(노드)**을 선택하고 이와 **지역적, 시간적**으로 연결된 관절 선택
 - 연결된 관절 선택 범위(D) 를 지정하는 함수를 Sampling function이라 정의
 - 소개하는 논문에서는 D 값을 1로만 사용

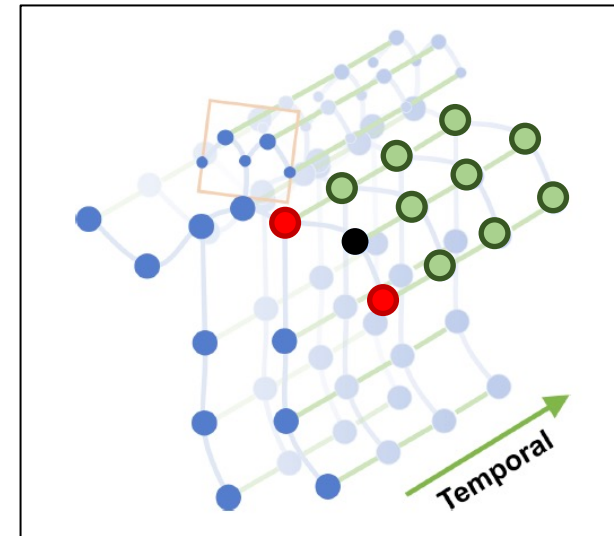
모든 연결이 진행된 그래프



특정 관절 선택



연결된 관절 선택(D=1)



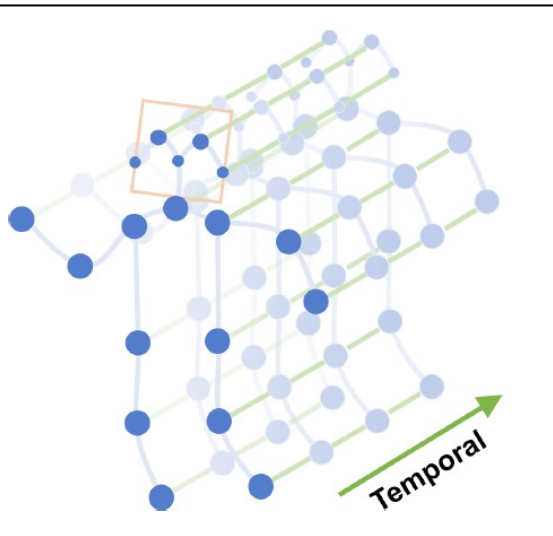
Spatial Temporal Graph Convolutional Networks

- ST-GCN

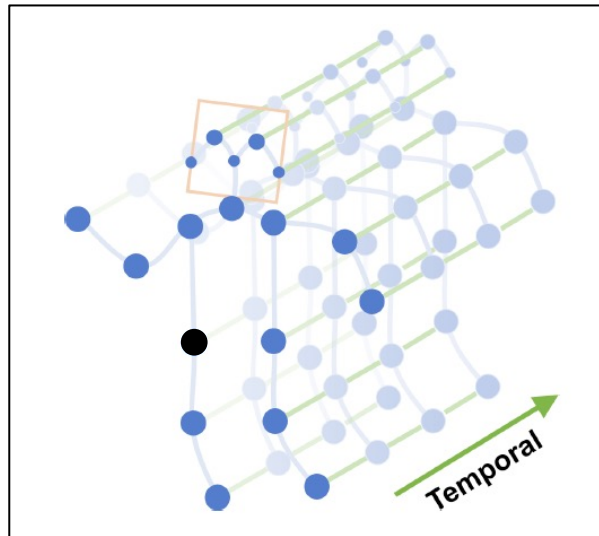
❖ Spatial Temporal Graph Convolutional Networks(ST-GCN)

- 개별 영상 내 모든 프레임에 대해 연결 진행(Intra-body and Inter-frame connection)
- 그래프 내 **특정 관절(노드)**을 선택하고 이와 **지역적, 시간적**으로 연결된 관절 선택
 - 연결된 관절 선택 범위(D) 를 지정하는 함수를 Sampling function이라 정의
 - 소개하는 논문에서는 D 값을 1로만 사용

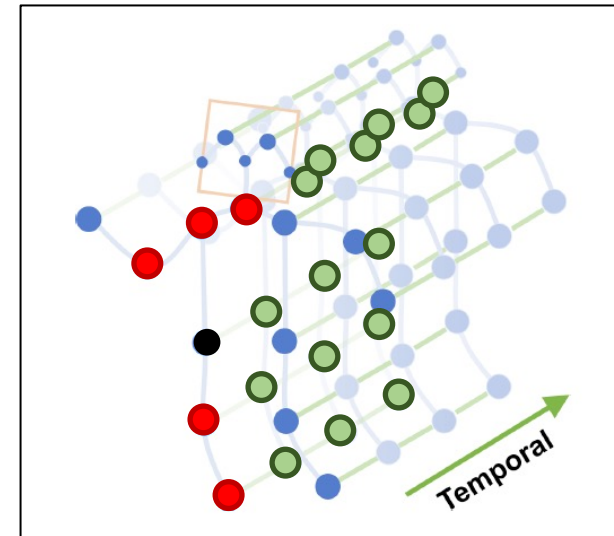
모든 연결이 진행된 그래프



특정 관절 선택



연결된 관절 선택(D=2)



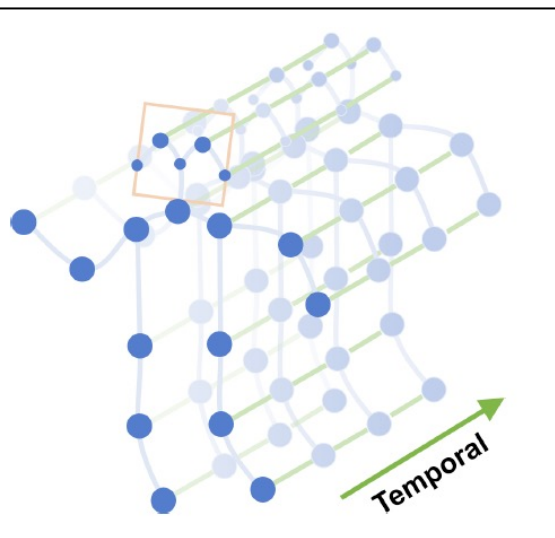
Spatial Temporal Graph Convolutional Networks

- ST-GCN

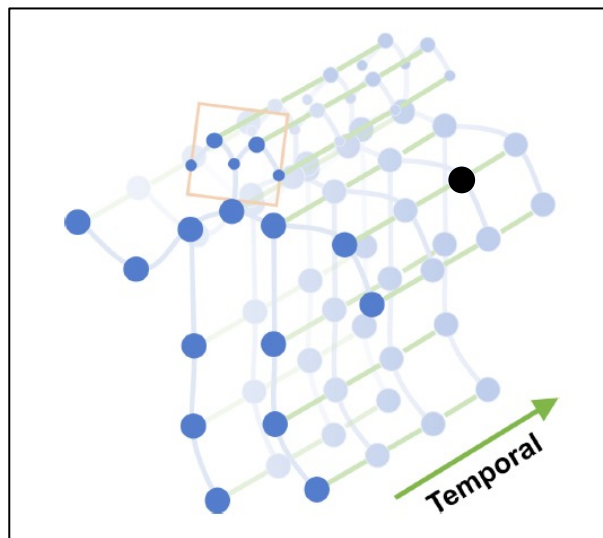
❖ Spatial Temporal Graph Convolutional Networks(ST-GCN)

- 모든 프레임에 대한 Skeleton을 사용해 모델 학습하는 것은 연산적에서 비효율
- 하이퍼파라미터 Γ 를 지정해 특정 시점 기준 과거 $\frac{\Gamma}{2}$ 시점과 미래 $\frac{\Gamma}{2}$ 시점에 대해 연산
 - Γ 가 2인 경우 과거 1시점, 미래 1시점 만 ST-GCN 연산 진행

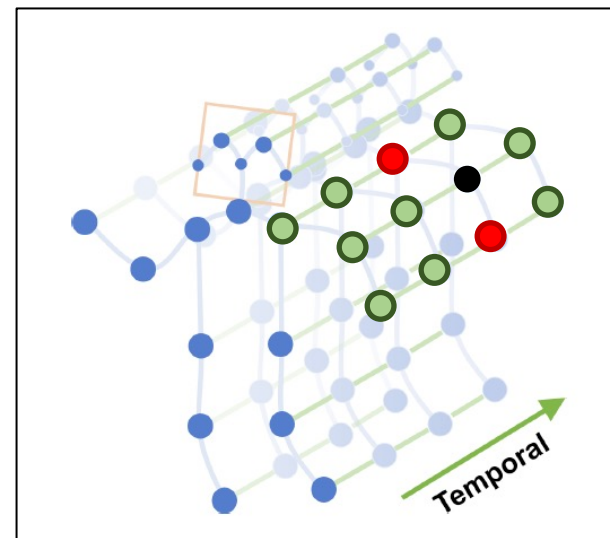
모든 연결이 진행된 그래프



특정 관절 선택



하이퍼파라미터 Γ 가 2인 경우

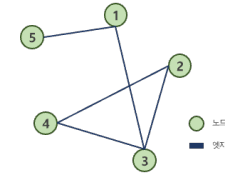


Spatial Temporal Graph Convolutional Networks

- ST-GCN

❖ Spatial Temporal Graph Convolutional Networks(ST-GCN)

- Graph Convolutional Networks(GCN) 연산 방식에 대한 설명
- 수식 내 약어에 대한 설명
 - f_{out} : GCN 연산 결과 | $\Lambda^{ii} = \sum_j (A^{ij} + I^{ij})$ | f_{in} : 관절(노드) 특징 벡터 | W : GCN 가중치
 - A : 인접 행렬 | I : 항등 행렬



인접 행렬
(Adjacency matrix)

	1	2	3	4	5
1			1		1
2					
3	1				
4					
5	1				

$$f_{out} = \Lambda^{-\frac{1}{2}} (A + I) \Lambda^{-\frac{1}{2}} f_{in} W$$

관절 특징 벡터

GCN 연산 결과

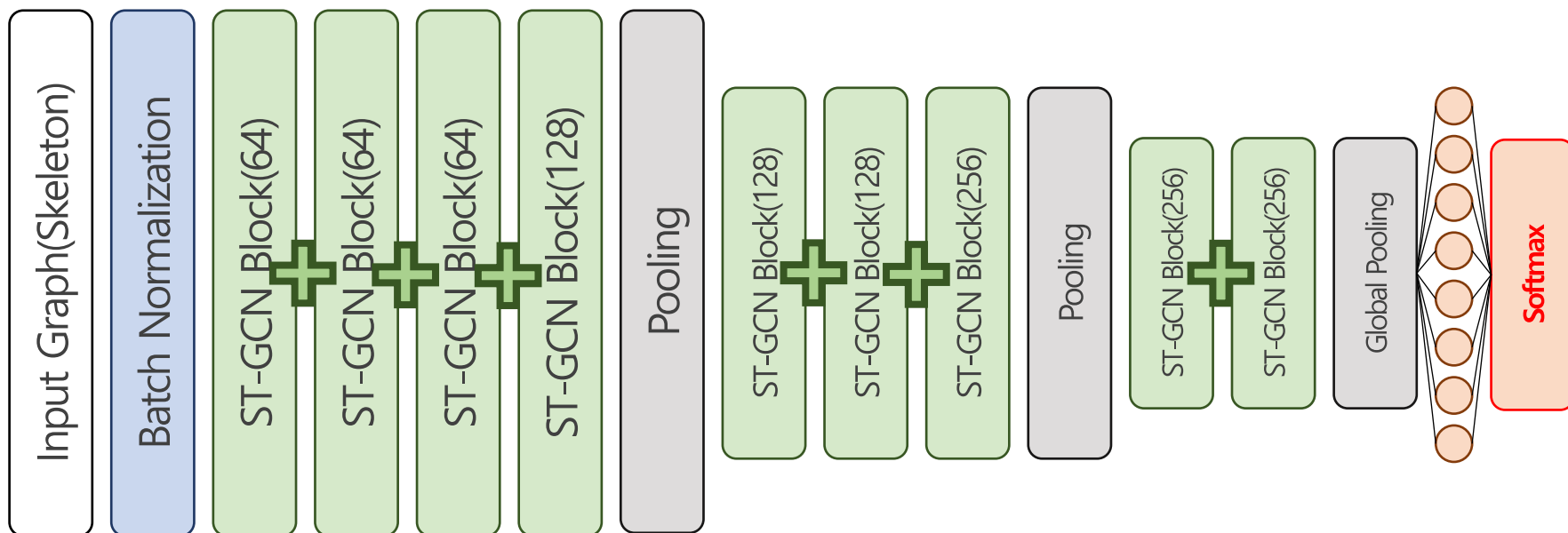
- 여러 가중치 채널(C)
- 그래프 내 선택된 관절 수(V)
- 선택된 시점 개수(T)

Spatial Temporal Graph Convolutional Networks

- Proposed model architecture

❖ Spatial Temporal Graph Convolutional Networks(ST-GCN)

- 논문에서 제안하는 ST-GCN 기반 행동 인식 모델
- **Batch normalization** 레이어는 Skeleton 크기 변화에도 일정한 성능을 보이기 위해 추가
 - Skeleton 크기를 일정한 수준으로 유지하기 위한 파라미터를 학습하기 위한 레이어
- **9개 ST-GCN 블록**으로 구성되며 ResNet과 같이 블록 사이 **잔차학습** 진행
- 마지막 연산은 **Softmax 연산**이며 행동 범주 별 확률 값 도출



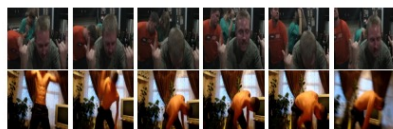
Spatial Temporal Graph Convolutional Networks

- Dataset

❖ 실험에 사용한 데이터셋

- Kinetics: YouTube에서 수집한 동작 영상 30만 클립과 이에 대한 행동(400) 분류 정보 존재
 - OpenPose를 사용해 2차원 관절 좌표와 관절별 확률 값 산출
- NTU-RGB+D: 5만 6천 행동 영상과 행동 종류 60개, 3차원 관절 좌표 정보 존재
 - RGB 영상과 원근감 인식이 가능한 Depth 카메라로 인간 행동 영상을 촬영
- 2D Skeleton(Kinetics), 3D Skeleton(NTU-RGB+D) 각각을 입력 데이터로 하는 ST-GCN 학습

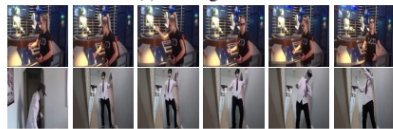
Kinetics



(a) headbanging



(c) shaking hands



(e) robot dancing



(b) stretching leg



(d) tickling



(f) salsa dancing

NTU-RGB+D



- Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., ... & Zisserman, A. (2017). The kinetics human action video dataset. arXiv preprint arXiv:1705.06950.

- Shahroudy, A., Liu, J., Ng, T. T., & Wang, G. (2016). Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1010-1019).

Spatial Temporal Graph Convolutional Networks

- Performance about Kinetics dataset

❖ Kinetics 데이터셋 에 대한 실험 결과

- 다양한 상황의 YouTube영상이 포함되어 있으며 Skeleton은 OpenPose 모델로 추출
- Top 1 정확도와 Top 5 정확도로 모델 성능 비교
- Video-based HAR 방법론, Skeleton-based 과거 방법론과 비교
 - 직접 정의한 관절 별 특징 벡터, 시간 변화에만 집중한 특징 벡터를 입력 변수로 사용
- Skeleton-based 과거 방법론보다는 성능이 높았지만 Video-based 방법론 보다는 낮은 성능
- Video-based 방법론 대비 성능이 낮은 이유는 영상과 행동 인식 범주 사이 부정확한 매칭

	Top-1	Top-5
RGB(Kay et al. 2017)	57.0%	77.3%
Optical Flow (Kay et al. 2017)	49.5%	71.9%
Feature Enc. (Fernando et al. 2015)	14.9%	25.8%
Deep LSTM (Shahroudy et al. 2016)	16.4%	35.3%
Temporal Conv. (Kim and Reiter 2017)	20.3%	40.0%
ST-GCN	30.7%	52.8%

Spatial Temporal Graph Convolutional Networks

- Performance about NTU-RGB+D

❖ NTU-RGB+D 데이터셋 에 대한 실험 결과

- 제한된 환경에서 촬영한 영상이며 영상 수집 동시에 3D 관절 좌표 수집
- X-Sub: 학습 데이터 내 행동하고 있는 인간이 검증 데이터 내에는 존재하지 않는 경우
- X-View: 동일 행동을 두 카메라를 사용해 촬영해 카메라를 기준으로 학습/검증 데이터 구분
- 두 가지 기준에 대한 성능을 Top 1 정확도를 사용해 비교
- 기존 최고 성능을 가진 모델 대비 큰 폭으로 성능 향상

	X-Sub	X-View
Lie Group (Veeriah, Zhuang, and Qi 2015)	50.1%	52.8%
H-RNN (Du, Wang, and Wang 2015)	59.1%	64.0%
Deep LSTM (Shahroudy et al. 2016)	60.7%	67.3%
PA-LSTM (Shahroudy et al. 2016)	62.9%	70.3%
ST-LSTM+TS (Liu et al. 2016)	69.2%	77.7%
Temporal Conv (Kim and Reiter 2017).	74.3%	83.1%
C-CNN + MTLN (Ke et al. 2017)	79.6%	84.8%
ST-GCN	81.5%	88.3%

Spatial Temporal Graph Convolutional Networks

- Summary of ST-GCN

❖ Spatial Temporal Graph Convolutional Networks(ST-GCN) 논문 요약

- 가장 기본적인 GCN 기반 Skeleton-based HAR 알고리즘
- 영상 데이터에 대해 OpenPose(Human pose estimation)를 사용해 Skeleton 추출
- 여러 프레임에 해당하는 Skeleton을 ST-GCN 입력 데이터로 사용
- ST-GCN을 사용해 Skeleton 특징 추출 및 행동 인식(분류) 진행



Input Video

목차

- ❖ Introduction
- ❖ Spatial-temporal Graph Convolutional Networks
- ❖ **Applications for Skeleton-based HAR**
- ❖ Conclusion

Applications for Skeleton-based HAR

- Skeleton-based HAR for Baseball

❖ Baseball Swing Pose Estimation Using OpenPose

- 야구 경기 중 배트로 날아오는 공을 맞추는 스윙 자세 평가를 위한 방법론 제시
- OpenPose 알고리즘을 사용해 스윙하고 있는 자세에 대한 관절 정보 파악
- 파악한 관절을 사용해 각종 거리, 팔 각도 등을 파악하고 이를 점수화
- 인공지능 알고리즘 중 규칙 기반(Rule-based) 알고리즘을 사용해 HAR 진행



OpenPose를 사용해 스윙 자세 인식



규칙	조건	점수
엉덩이 관절 사이 거리	거리 < 76.8 pixels	30점
왼쪽 팔꿈치 각도	$145^{\circ} < \theta < 180^{\circ}$	30점
오른쪽 팔꿈치 각도	$100^{\circ} < \theta < 130^{\circ}$	20점
왼발 무릎 각도	$150^{\circ} < \theta < 170^{\circ}$	20점
오른발 무릎 각도	$115^{\circ} < \theta < 135^{\circ}$	15점
어깨 각도	$170^{\circ} < \theta < 190^{\circ}$	15점

- Li, Y. C., Chang, C. T., Cheng, C. C., & Huang, Y. L. (2021, April). Baseball Swing Pose Estimation Using OpenPose. In 2021 IEEE International Conference on Robotics, Automation and Artificial Intelligence (RAAI) (pp. 6-9). IEEE.

Applications for Skeleton-based HAR

- Skeleton-based HAR for Cyber Physical Systems

❖ Cyber Physical Systems with Human in the Loop

- Cyber physical system이 인간 행동을 인식하고 행동에 대한 결과를 제시
- 인간이 제시한 결과를 확인하고 다른 행동을 진행하며 둘 사이에 소통 진행
- 인간 행동 인식을 위해 Skeleton을 추출하고 주성분 분석을 사용해 차원 축소
- 차원 축소된 Skeleton 특징 벡터를 입력해 위험 여부 분류 모델 학습



타격
승구
수비
.....

Smart Factory



Cyber Physical
Systems

Smart Factory 내 직원



Skeleton 및
PCA 특징 벡터
추출

좋은 자세

자세 교정
진행

위험한 자세

- Nikolov, P., Boumbarov, O., Manolova, A., Tonchev, K., & Poulkov, V. (2018, July). Skeleton-based human activity recognition by spatio-temporal representation and convolutional neural networks with application to cyber physical systems with human in the loop. In 2018 41st International Conference on Telecommunications and Signal Processing (TSP) (pp.1-5). IEEE.

Applications for Skeleton-based HAR

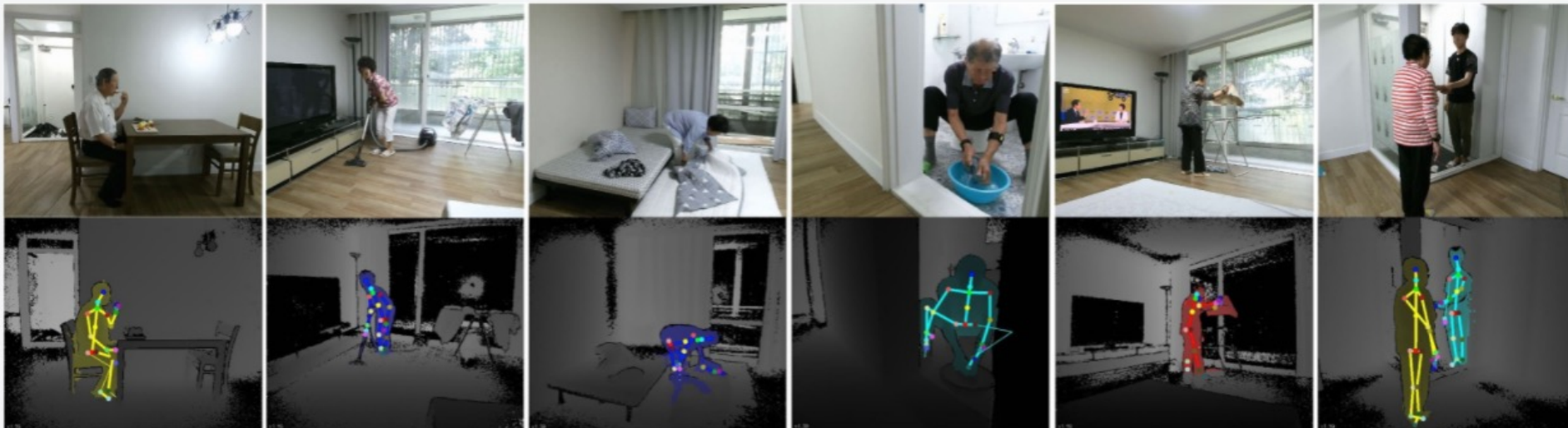
- Skeleton-based HAR for Human Care System

❖ 고령자 일상 행동 인식을 위한 휴먼 케어 로봇 개발

- 대한민국 고령화 속도가 가속되면서 독거 노인 가구 급증
- 독거 노인을 위한 휴먼 케어 로봇이 개발되었고 행동 인식 기술이 로봇 기술 내 포함
- 일상적인 행동 뿐만 아니라 쓰러지는 상황 같이 위험한 행동 범주를 포함한 데이터 공개
- RGB 영상과 원근감 인식을 위한 Depth 영상이 존재하며 이에 대한 3차원 좌표 정보 존재



타격
송구
수비
.....



- Jang, J., Kim, D., Park, C., Jang, M., Lee, J., & Kim, J. (2020, October). ETRI-activity3D: A large-scale RGB-D dataset for robots to recognize daily activities of the elderly. In 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 10990-10997). IEEE

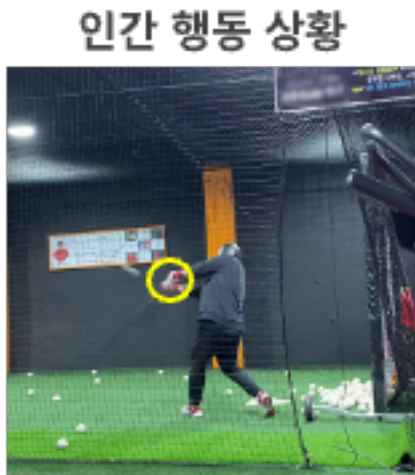
목차

- ❖ Introduction
- ❖ Spatial-temporal Graph Convolutional Networks
- ❖ Applications for Skeleton-based HAR
- ❖ **Conclusion**

Conclusion

❖ Human Activity Recognition(HAR) 요약

- **시그널** 또는 **영상**을 입력 받아 **인간이 어떠한 행동을 하고 있는 지를 인식하는 문제**



시그널 정보 수집



영상 정보 수집

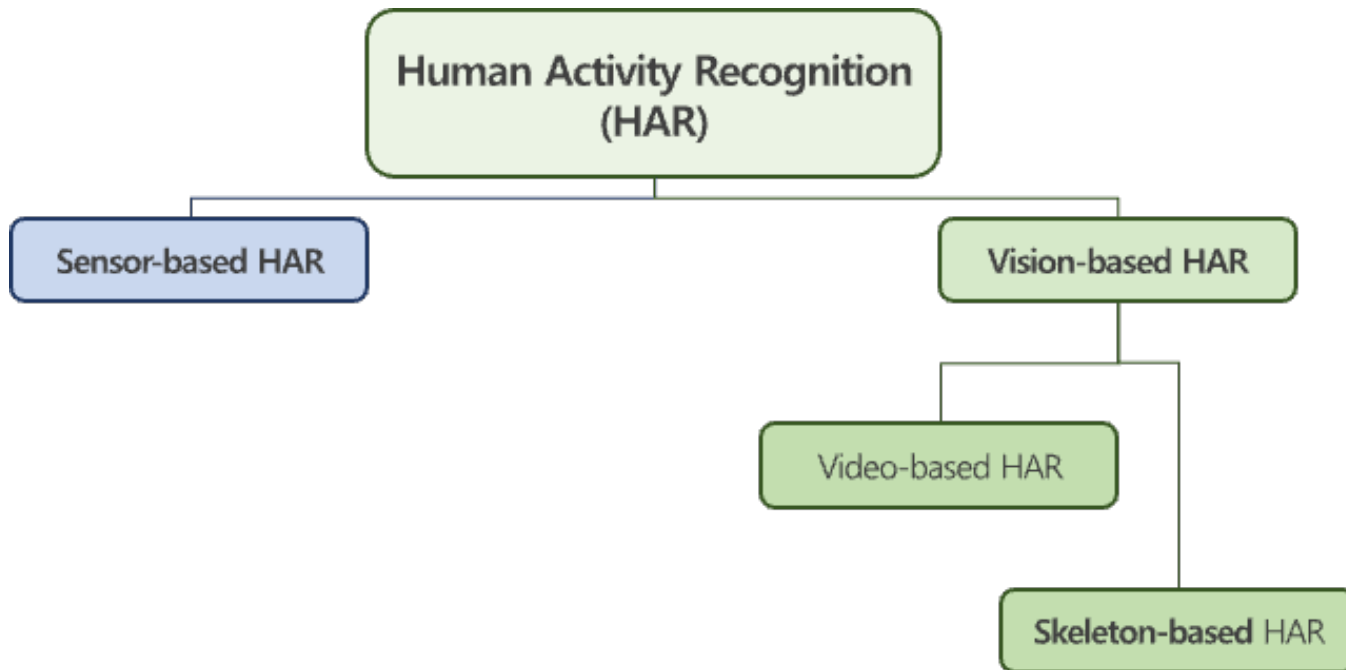


타격
송구
수비
.....

Conclusion

❖ Human Activity Recognition(HAR) 요약

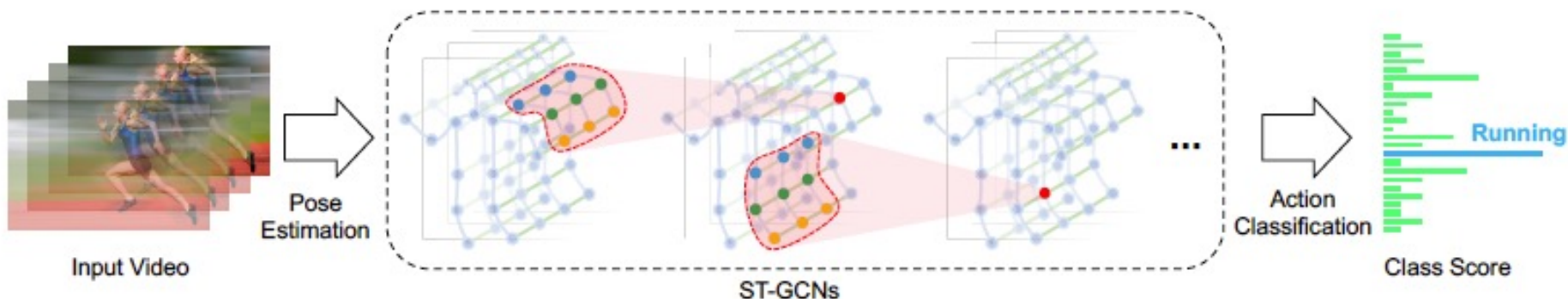
- **시그널** 또는 **영상**을 입력 받아 **인간이 어떠한 행동**을 하고 있는 지를 **인식**하는 문제
- **시그널 데이터**를 입력 받아 행동 인식하는 알고리즘은 Sensor-based HAR
- **영상 데이터**를 입력 받아 행동 인식하는 알고리즘은 Vision-based HAR
 - 영상 데이터 사용 방식에 따라 Video-based와 Skeleton-based 로 구분 가능



Conclusion

❖ Spatial Temporal Graph Convolutional Networks(ST-GCN) 요약

- 입력 데이터는 영상에서 추출된 Skeleton(Human pose estimation 결과)
- 출력 데이터는 Skeleton이 움직이고 있는 행동(범주 정보)
- Graph Convolutional Networks를 사용해 특징 추출하고 최종적으로 분류 진행



❖ Skeleton-based HAR 응용 분야

- 야구와 같은 자세가 중요한 스포츠에서 인공지능 기반 자세 평가 자동화에 사용
- CCTV 영상을 입력 받아 인간 행동을 인식하고 이상 행동 시 알람을 주는 시스템 개발
- 독거 노인을 위한 휴먼 케어 로봇 개발에 인식 기술로 탑재

감사합니다.