

골프 클럽과 인간 관절의 통합: 스윙 분석을 위한 시공간 그래프 합성곱 기법

정승현¹, 김민석¹, 김현진¹, 홍서영², 최재림², 이우진^{3,*}

¹동국대학교 컴퓨터·AI학과 석사과정

²(주)김캐디

³동국대학교 컴퓨터·AI학과 조교수

alzkdpf23@dgu.ac.kr, cdi6578296@dgu.ac.kr, tkrhk8011@dgu.ac.kr,

seoyoung@kimcaddie.com, jrim@kimcaddie.com, wj926@dgu.ac.kr

Integrating Golf Club and Human Joints: A Spatial-Temporal Graph Convolutional Method for Swing Analysis

Seunghyeon Jung¹, Minseok Kim¹, Hyeonjin Kim¹, Seoyoung Hong², Jaerim Choi², Woojin Lee^{3,*}

^{1,3}Dept. of Computer Science and Artificial Intelligence, Dongguk University

²Kimcaddie Inc

요약

골프는 정밀한 신체 움직임을 요구하는 스포츠로, 미묘한 스윙의 변화가 공의 궤적에 상당한 영향을 미칠 수 있다. 본 연구는 다양한 골퍼의 스윙 영상과 공 궤적 정보를 직접 수집한다. 이를 통해 클럽의 위치 정보와 인체 관절 정보를 통합하는 새로운 골프 스윙 분석 방법론을 제안한다. 시공간 그래프 합성곱 기법을 사용하여 스윙 자세와 공 궤적 간의 관계를 분석한다.

1. 서론

골프는 정밀한 신체 움직임을 제어할 필요가 있는 스포츠로, 스윙에서의 미묘한 변화를 포착하고 공의 궤적에 미치는 영향을 분석하는 것이 중요하다. 특히 AI의 발전으로 다양한 방식의 스윙 분석 연구가 진행되고 있다. 그러나 기존의 연구는 클럽의 영향을 고려하지 못하여 골프라는 스포츠의 특성을 반영하지 못하였고, 공 궤적 정보가 부족하여 스윙의 결과를 제대로 반영하지 못하는 한계가 존재한다 [1, 2, 3]. 본 연구는 스윙 영상과 해당하는 공 궤적 정보를 동시에 수집하여 이러한 문제를 해결한다. 또한, 클럽의 위치 정보를 추적하고 이를 인체 관절 정보와 연결함으로써 그래프 구조의 데이터를 구축한다. 시공간 그래프 합성곱 기법을 사용하여 스윙 자세와 공 궤적 간의 관계를 분석한다.

2. 관련연구

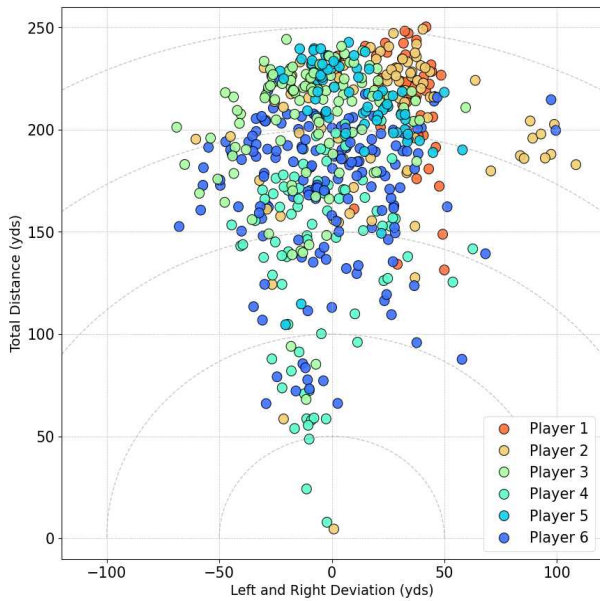
인체 동작 인식 (Human Action Recognition, HAR)은 영상 데이터로부터 인간의 행동을 이해하고 해석하는 데 중점을 둔 컴퓨터 비전의 핵심 연구 분야이다. 그중 인체 관절 정보 기반 동작 인식은

동적이고 복잡한 환경에서 견고한 성능을 보인다. 원본 RGB 영상에 의존하는 전통적인 접근 방식과 달리, 관절 정보 기반 표현은 배경 간섭, 조명과 같은 잡음에 대해 강건하다 [4].

그래프 합성곱 신경망 (Graph Convolutional Networks, GCNs)은 인체 관절 정보가 이루는 구조적 정보를 포착하는 데 탁월한 성능을 보여준다. 이중 시공간 그래프 합성곱 기법 (Spatial-Temporal Graph Convolutional Network, ST-GCN)은 관절 정보를 노드(nodes)로, 관절 간의 연결을 엣지(edges)로 하는 그래프 구조로 자연스럽게 표현하므로 인체 동작 인식에 매우 효과적이다 [5].

2. 데이터 수집

본 연구의 데이터 수집은 (주)김캐디와의 협업을 통해 진행한다. 여섯 명의 참가자로부터 스윙 영상과 공 궤적 정보를 직접 수집한다. 참가자들은 모두 아마추어 남성 여섯 명으로 구성되며, 외부 변수를 최소화하기 위해 통제된 실내 스크린골프 연습장에서 수집을 진행한다. 모든 참가자는 동일한 두 개의 드라이버를 사용하여 스윙하며, 가능한 공을 멀리 그리고 정확하게 치도록 지시하였다.



(그림 1) 수집한 공 궤적 분포

공 궤적 정보는 TrackMan4 장비를 활용하여 자동으로 계산된 정보를 수집한다. 총 563개의 스윙 영상과 공 궤적을 포함한 데이터셋을 수집하였고, (그림 1)은 참가자에 따른 공 궤적 분포를 나타낸다.

3. 방법론

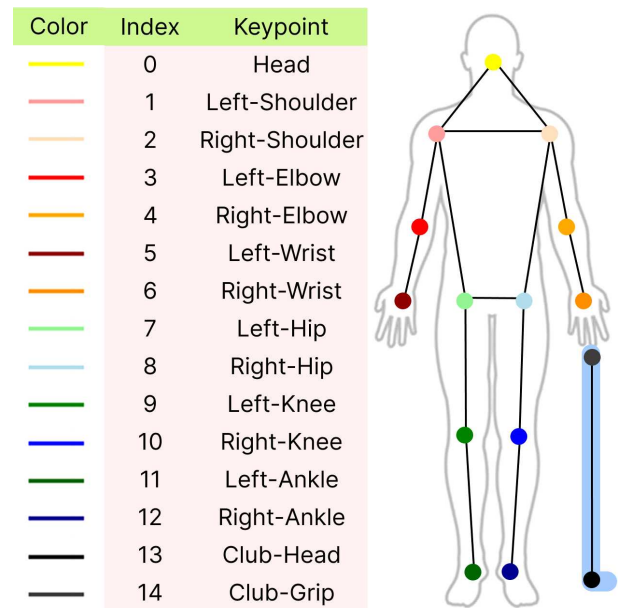
3.1 클럽 위치 정보 추적

골프 스윙 분석의 고유한 특성을 반영하기 위해 클럽의 위치 정보를 추적한다. 골프는 클럽을 사용하는 스포츠이므로 공을 타격하는 클럽 헤드와 골퍼의 몸에 연결되는 클럽 그림의 위치 정보는 중요한 키포인트 역할을 한다. 그러나 스윙 중 클럽은 매우 빠른 속도로 움직이기 때문에, 기존의 YOLO 모델로는 이러한 움직임을 정확하게 포착하기 어려울 수 있다. 본 연구에서는 수집된 스윙 영상으로부터 클럽 헤드와 클럽 그림의 정보를 미세 조정하여 문제를 해결한다. 먼저, 클럽 헤드와 클럽 그림에 대한 바운딩 박스를 라벨링 한다. 이어서, 객체 탐지를 수행하기 위해 사전 훈련된 YOLO 모델을 미세 조정한다.

3.2 인체 관절 정보 수집 및 그래프 구조 통합

본 연구는 골퍼의 움직임에서 인체 관절 정보를 추출하기 위해 YOLO-pose 모델을 사용한다. 이 모델은 총 17개의 인체 관절 정보를 추출하며 두 개의 눈, 두 개의 귀, 그리고 하나의 코 정보를 포함한다. 그러나 카메라 각도로 인해 발생하는 가려짐

(occlusion)으로 인해 눈과 귀의 정확도가 낮게 나타난다. 코 정보만으로도 스윙 분석을 위한 머리의 움직임을 포착할 만하다고 판단하여 분석에서 눈과 귀 관절 정보를 제거한다. 이에 3.1절에서 추적한 클럽 헤드와 클럽 그림 정보를 통합하여 그래프 구조를 만든다. (그림 2)는 본 연구에서 사용하는 15개의 키포인트의 이름과 그래프 구조를 나타낸다. 이를 통해 클럽의 움직임과 인체 관절의 움직임을 모두 포착하는 스윙 분석을 진행할 수 있다.



(그림 2) 클럽 정보와 관절 정보를 통합한 그래프 구조

4. 실험

본 연구는 제안한 방법론의 성능을 기존 연구 접근 방식과 비교한다. 본 연구에서 제안한 그래프 구조의 특성을 잘 반영하기 위해 ST-GCN 모델을 사용한다. ST-GCN은 스윙의 공간적, 시간적 정보를 모두 잘 반영할 수 있는 모델이다. 또한, ViT, ResNet, LSTM, RNN과 같은 다양한 모델을 비교 모델로 사용한다 [6, 7, 8, 9]. ViT와 ResNet은 스윙의 공간적 정보를 잘 반영할 수 있는 비교 모델이고, LSTM과 RNN은 스윙의 시간적 정보를 잘 반영할 수 있으므로 비교 모델로 사용한다. 모든 모델은 공정하고 신뢰할 수 있는 비교를 보장하기 위해 동일한 데이터 전처리 기술을 활용하여 일관된 실험 조건에서 훈련 및 평가한다.

4.1 데이터 전처리

2절에서 수집한 데이터셋에서 이상치를 제거하여

데이터를 전처리한다. 공의 비거리가 150야드 미만인 데이터와 좌우로 30야드 이상 벗어나는 데이터는 제거한다.

4.2 실험 환경

공정한 비교를 위해 모든 모델은 동일한 조건에서 학습한다. 옵티마이저는 AdamW을 사용하고, 학습률은 $1E-3$ 에서 $1E-5$ 의 범위 내에서 적응형 학습률을 사용한다. 배치 크기는 16으로 고정되며, MSE와 Cross-Entropy 손실 함수를 적용한다. 모든 실험은 단일 NVIDIA RTX 4090 GPU에서 수행한다.

4.3 실험 결과

수집한 공 궤적 정보 중 3가지 변수에 대한 실험을 진행한다. 골프 스윙의 주요 목표는 볼을 정확하게 멀리 보내는 것이다. **스핀축** 변수와 **좌우방향각** 변수는 공을 정확하게 보내는 목표를 반영하는 변수이고, **스피드** 변수는 공을 멀리 보내는 목표를 반영하는 변수이다.

스핀축 변수는 비행 중 볼이 얼마나 휘는지를 측정한다. 스핀축 변수는 -5 도에서 5 도 사이의 임계값을 사용하여 공 궤적을 이진 분류한다. 마찬가지로 공의 초기 각도를 나타내는 좌우 방향각 변수는 -3 도에서 3 도 사이의 임계값을 사용하여 이진 분류한다. 스피드 변수는 골퍼가 더 긴 샷을 위해 파워를 생성하는 능력을 측정하는 데 효과적인 지표이다. 스피드 변수는 연속 변수이므로 회귀 접근 방식을 채택한다. 이러한 방식으로 예측 모델을 구성함으로써, 골프 스윙의 주요 성능 측면을 평가하고 데이터셋의 타당성을 검증한다.

<표 1> 실험 결과

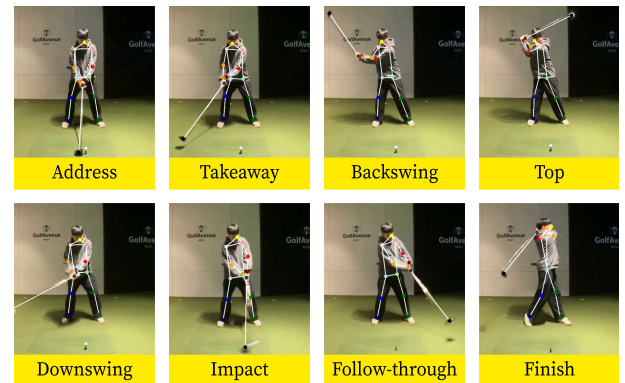
모델	스핀축		좌우방향각		스피드	
	AUC	정확도	AUC	정확도	R^2	RMSE
ViT	0.7950	0.5714	0.7904	0.5932	0.7560	6.7117
ResNet	0.7805	0.6886	0.6444	0.6275	0.7768	6.3232
LSTM	0.7791	0.6275	0.7572	0.8281	-	-
RNN	0.7784	0.6275	0.8299	0.8281	-	-
Ours*	0.8294	0.7297	0.8762	0.8333	0.9095	3.9692
Ours	0.8557	0.7297	0.9702	0.8750	0.9159	3.8254

<표 1>은 스핀축과 좌우방향각 변수에 대한 이진 분류 실험과 스피드 변수에 대한 회귀 실험 결과를 나타낸다. 여러 모델과의 비교를 통해 ST-GCN 모델을 활용한 본 연구의 방법론이 가장 우수한 성능을 나타냄을 보여준다. 특히, *로 표시한 실험은 클

럽 정보를 사용하지 않은 결과이다. 이를 통해 클럽 정보를 사용하는 것이 골프 스윙 분석에 큰 영향을 미치는 것을 확인한다.

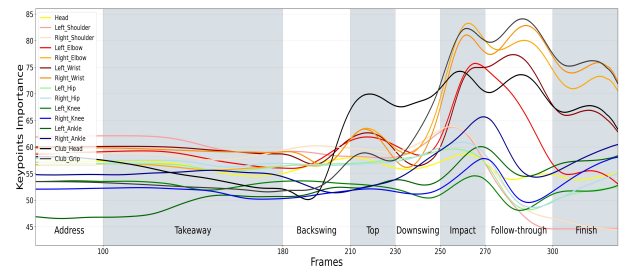
4.4 키포인트 분석

ST-GCN 모델은 각 키포인트의 중요도를 학습하므로 다양한 스윙 단계에 각 키포인트의 영향을 분석할 수 있다. 분석의 용이성을 위해 연속적인 스윙 동작을 어드레스, 테이크어웨이, 백스윙, 탑, 다운스윙, 임팩트, 팔로스루, 피니시의 8가지 단계로 분류한다 [10]. (그림 3)은 각 스윙 단계를 보여주며, 각 단계별 핵심 동작의 예시이다.



(그림 3) 스윙 단계별 핵심 동작

스피드 변수를 분석한 결과는 (그림 4)와 같다. 이를 통해 하체의 체중 이동을 통한 파워 생성과 임팩트 시 강한 공의 속도를 만들어내는 데 클럽의 영향력이 중요함을 확인할 수 있다.



(그림 4) 스피드 변수에 대한 키포인트 영향 분석

5. 결론

본 연구는 클럽 정보와 인체 관절 정보를 통합하여 스윙 분석의 새로운 접근법을 제안한다. 시공간 그래프 합성곱 기법을 통한 실험 결과는 다른 기법 대비 우수한 성능을 보이며, 개별적인 키포인트의 영향력에 대한 분석도 제공한다.

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업(IITP-2025-RS-2020-II201789)과 인공지능융합혁신인재양성사업(IITP-2025-RS-2023-00254592), 한국연구재단의 지원(No. RS-2025-00556289)으로 수행되었음.

참고문헌

- [1] Ju, Chan-Yang, Jong-Hyeon Kim, and Dong-Ho Lee. "GolfMate: enhanced golf swing analysis tool through pose refinement network and explainable golf swing embedding for self-training." *Applied Sciences* 13.20 (2023): 11227.
- [2] Kim, Theodore T., Mohamed A. Zohdy, and Michael P. Barker. "Applying pose estimation to predict amateur golf swing performance using edge processing." *IEEE Access* 8 (2020): 143769-143776.
- [3] Liao, Chen-Chieh, Dong-Hyun Hwang, and Hideki Koike. "Ai golf: Golf swing analysis tool for self-training." *IEEE Access* 10 (2022): 106286-106295.
- [4] Sun, Zehua, et al. "Human action recognition from various data modalities: A review." *IEEE transactions on pattern analysis and machine intelligence* 45.3 (2022): 3200-3225.
- [5] Yan, Sijie, Yuanjun Xiong, and Dahua Lin. "Spatial temporal graph convolutional networks for skeleton-based action recognition." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 32. No. 1. 2018.
- [6] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).
- [7] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [8] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.
- [9] Rumelhart, David E., Geoffrey E. Hinton, and

Ronald J. Williams. "Learning representations by back-propagating errors." *nature* 323.6088 (1986): 533-536.

- [10] McNally, William, et al. "GolfdB: A video database for golf swing sequencing." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2019.