스포츠 대상 실시간 영상분석 자세교정 및 피드백 모바일 어플리케이션

임승택, 장우진, 황준현, 손주완 한기준*

한성대학교 컴퓨터공학부

{2071334, keejun.han*}@hansung.ac.kr

Real-Time Sports Video Analysis, Posture Correction, and Feedback Application

Seungtaek Lim, Woojin Jang, Junhyeon Hwang, Juwan Son and Keejun Han*

Hansung Univ., *Hansung Univ.

요 약

최근 멀티모달 인공지능 모델의 발전으로 이미지와 텍스트를 동시에 이해하고 분석하는 기술이 다양한 분야에서 활용되고 있다. 그러나 이러한 범용 모델들은 여전히 도메인 특화 학습(Domain Adaptation) 과 실시간 응용(Real-Time Application) 측면에서 한계를 가진다. 특히 스포츠와 같이 동작 기반의 정밀한 시공간 분석이 필요한 분야에서는 실시간 응답성과 정확도를 동시에 만족하기 어렵다. 본 연구에서는 이러한 한계를 해결하기 위해 실내 스포츠 환경을 위한 실시간 자세 분석 및 피드백 모바일 시스템을 제안하였다. 제안 시스템은 온디바이스에서 자동 촬영 및 객체 감지(Object Detection) 를 수행하여 사용자의 동작을 실시간으로 포착하고, 촬영된 영상을 서버로 전송해 자세 추정(Pose Estimation) 을 수행한다. 이후 추출된 관절 데이터를 기반으로 주요 지표(어깨 각도 편차, 상체 이동 거리, 손목 이동 거리, 발목 변화 횟수, 사용자 숙련도 등)를 계산하고, 언어 모델을 활용한 자연어 피드백으로 사용자에게 직관적인 운동 교정 정보를 제공한다. 모델은 INT8 양자화를 통해 모바일 환경에 최적화되었으며, 실험 결과 추론 속도 74% 단축, 양자화로 인한 메모리 사용량 감소를 달성하였다. 본 연구는 전문 장비 없이도 실시간 분석과 자연어 피드백을 동시에 제공할 수 있는 구조를 제시함으로써, 스포츠 자세 교정, 반복 학습, 개인 맞춤형 피드백 제공 측면에서 높은 실용성과 확장 가능성을 지닌다.

Ⅰ. 서 론

최근 Vision-Language Models(VLM)과 같은 멀티모달 인공지능 모델은 이미지와 텍스트를 동시에 이해하며 다양한 응용 분야에서 활용되고 있다. 대표적으로 GPT-4o[1], DeepSeek-VL[2], Qwen 3-VL[3] 등이 있으며, 이들은 이미지 인식과 자연어 이해를 통합하여 질문 응답, 캡셔닝, 요약 등 복합적 태스크를 처리할 수 있다.

그러나 이러한 범용 모델들은 도메인 특화 학습과 실시간 응용 측면에서 여전히 한계를 가진다. 특히 스포츠나 헬스케어처럼 동작 기반의 정밀한 시공간 분석이 필요한 분야에서는 실시간 응답성과 정확도를 모두 만족하기 어렵다. 또한 개인별 피드백을 위해서는 프라이버시 보호와 온디바이스-서버 간 분산 처리 구조가 필수적이다.

이에 본 연구는 이러한 한계를 보완하기 위해 실내 스포츠 환경을 위한 실시간 자세 분석 및 피드백 모바일 시스템을 제안한다. 제안 시스템은 온디바이스 자동 촬영 및 객체 감지(Object Detection) 와 서버 기반 자세 추정(Pose Estimation), 그리고 언어 모델을 이용한 자연어 피드백 생성을 결합하여 실시간성, 보안성, 개인화 피드백을 동시에 달성한다.

Ⅱ. 시스템 개요

본 논문에서 모바일 실시간 영상분석 및 피드백 생성 시스템은 그림 1. 과 같이 하이브리드(On-Device + Server) 구조로 설계되었다.

1) On-Device 단계:
모바일 디바이스에서는 YOLOv11 Object Detection을 수행하여 사용자의 투구 동작을 자동으로 감지한다. 이를 통해 사용자의 수동 조작 없이 자동 촬영 타이밍을 제어한다.

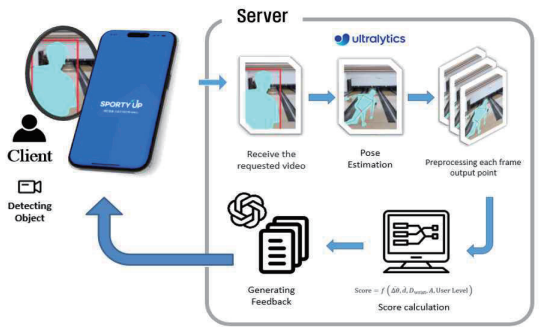

그림 1. 스포츠 자세 분석을 위한 하이브리드 시스템 아키텍처
Fig. 1. Hybrid system architecture for posture analysis

2) Server 단계:
촬영된 영상은 보안 전송 과정을 거쳐 Flask 기반 분석 서버로 전송된다. 서버의 YOLOv11 Pose Estimation[4] 모델은 각 프레임마다 17개의 신체 관절 좌표를 추출하며, 추출된 데이터를 기반으로 어깨 평균 각도 편차($\overline{\Delta\theta}$), 상체 평균 이동 거리($\bar{d}$), 손목 누적 이동 거리(D_{wrist}), 발목 위치 변화 횟수(N_{change}), 사용자 정보를 통해 습득한 사용자의 숙련도(User Level) 지표를 계산하고, 이를 다음과 같은 점수 함수로 종합하여 하나의 스코어로 환산한다.

$$Score = f(\overline{\Delta\theta}, \bar{d}, D_{wrist}, N_{change}, UserLevel) \quad (1)$$

스코어는 EXCELLENT, GOOD, COMMON, BAD의 네 등급으로 분류되며, 각 지표 값은 ChatGPT-4 API의 시스템 프롬프트로 전달되어 사용자에게 자연어 형태의 피드백 문장으로 제공된다.

모바일 환경은 연산 자원, 메모리 용량, 전력 소비 등의 제약이 존재하므로 딥러닝 모델의 경량화가 필수적이다.

본 시스템에서는 YOLOv11 Object Detection 모델을 Qualcomm AI Hub를 활용하여 weights와 activation을 각각 8비트 고정소수점 및 INT8 양자화(W8A8) 하였고, TensorFlow Lite 형식으로 변환해 모바일 환경에 최적화하였다. 실험은 Qualcomm Snapdragon 8 Elite 기반 디바이스(Samsung Galaxy S25) 에서 수행되었으며, 최적화 전후의 성능 비교는 다음 그림 2.와 같다.

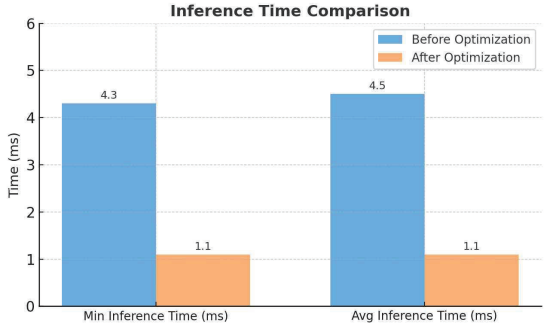

그림 2. 모델 최적화 전후의 성능 비교

Fig. 2. Performance comparison before and after model optimization

최적화 결과, 추론 시간은 약 74% 단축, 그리고 양자화를 통해 메모리 사용량은 감소하였으며, 모바일 디바이스에서도 빠른 응답성과 낮은 자원 소모를 동시에 만족시키는 성능을 확인할 수 있었다.

Ⅲ. 결론

본 연구에서는 Vision-Language Model이 보편화된 시대에, 범용 VLM이 해결하지 못한 도메인 특화 및 실시간 피드백 문제를 해결하기 위한 실내 스포츠 자세 분석 시스템을 제안하였다. 온디바이스의 자동 촬영 및 객체 감지, 서버의 자세 추정 및 점수 산출, LLM을 이용한 자연어 피드백 생성을 결합하여 실시간성과 정확도를 모두 확보하였다.

또한 INT8 양자화를 통해 경량화된 모델 구조와 실시간 처리 속도를 달성함으로써, 실제 모바일 환경에서도 효과적인 추론이 가능함을 입증하였다. 향후에는 본 시스템을 기반으로 골프, 요가, 탁구 등 다양한 종목으로 확장하고, 비전-언어-행동(Vision-Language-Action) 융합 기반의 지능형 운동 피드백 시스템으로 발전시킬 예정이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 2024년도 SW중심대학사업의 결과로 수행되었음(2024-0-00049)

참 고 문 헌

[1] OpenAI, J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, et al., “GPT-4 Technical Report,” arXiv preprint, arXiv:2303.08774, Mar. 2023.

[2] H. Lu, W. Liu, B. Zhang, B. Wang, et al., “DeepSeek-VL: Towards Real-World Vision-Language Understanding,” arXiv preprint, arXiv:2403.05525, 2024.

[3] A. Yang, A. Li, B. Yang, B. Zhang, B. Hui, et al., “Qwen3 Technical Report,” arXiv preprint, arXiv:2505.09388, 2025.

[4] Maji, D., Nagori, S., Mathew, M., & Poddar, D., "YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss," arXiv preprint arXiv:2204.06806, 2022.