International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Survey of Computer Vision Techniques for Multi-Athlete Detection, Pose Tracking and Real-Time Coaching Support

Guide: Prof. Aboli More
Ishwari Shivaji Kadam
Prajakta Nitin Shirsat
Siddhi Sameer Khambe
Gayatri Santram Sathe
Payal Mohan Sulkar

Department of Artificial Intelligence and Machine Learning,
Zeal Polytechnic, Nhare, Pune, Maharashtra, India

ABSTRACT

A coaching system that uses a camera needs four main parts: finding people in the video, tracking each person, finding body points (pose), and checking sports movements. This paper reviews common methods for each part and compares their strengths and limits for real-time use with many athletes. We also look at existing sports analysis systems and datasets, list open problems such as blocked views, ID mix-ups, privacy and sport-specific rules, and suggest future work. At the end, we explain what this study means for Athletix AI, our proposed coaching system that is still being built. This is a review of published work. It has no new experiments, and all comparisons are descriptive and should be checked in the cited papers.

KeyWords —*survey, pose estimation, multi-object tracking, person detection, sports analytics, real-time coaching.*

I. INTRODUCTION

A coach often trains many athletes at the same time and cannot note every mistake during practice. Feedback on technique still depends mainly on the coach's eyes or on costly capture systems. With deep learning, a normal camera can now find people, track them and estimate their body pose almost in real time [1], [2]. This makes a low-cost, camera-only coaching tool possible.

Building such a tool needs many choices, and the research is spread over several fields: detection, tracking, pose estimation, action recognition and sports analytics. This survey puts them together for one goal: a live-camera assistant that handles many athletes, keeps track of each one, and reports their movements to the coach.

The main points of this survey are: (i) a simple classification of the techniques needed (Fig. 1); (ii) a comparison of well-known methods; (iii) a review of sports systems and datasets; (iv) open problems and future directions; and (v) what this means for the design of Athletix AI.

II. SCOPE OF THE SURVEY

This is a narrative review. We chose well-known and widely cited papers that are useful for real-time, multi-person analysis of athletes from a camera. We cover only 2D, single-camera, normal (RGB) video. Wearable sensors, marker-based motion capture, 3D pose and multi-camera setups are not covered. [Authors: if you do a systematic search, write the databases, keywords, years and selection rules here.]



Fig. 1. Classification of the techniques reviewed in this paper.

III. BACKGROUND

A camera-based coaching system usually works in this order: capture video → detect people → track them across frames → find body points (pose) → calculate features such as joint angles → check the sport rules → give a report. Detection finds each person in a frame. Tracking connects the same person across frames and gives a temporary ID. Pose estimation finds points such as shoulders, elbows, hips and knees. Analysis compares these features with sport-specific rules. A mistake in an early step affects all later steps, so the weakest part often limits the whole system.

The angle at a joint B, formed by points A, B and C, is calculated as $\theta = \arccos((BA \cdot BC) / (\|BA\| \|BC\|))$. Joint angles are simple features and easy for a coach to understand.

IV. PERSON DETECTION

Two-stage detectors such as Faster R-CNN first suggest regions and then classify them. They are usually accurate but slower. Single-stage detectors such as the YOLO family find people in one pass, so they are fast and suit live video. For sports, a detector must work well for small, far, fast and partly hidden people. YOLO gives a good balance of speed and accuracy and has many ready tools, but crowded scenes and very small people are still difficult, while Faster R-CNN is not ideal for low-power live use.

V. MULTI-PERSON TRACKING

Most real systems first detect people in each frame and then link them over time. SORT uses a Kalman filter to predict motion and the Hungarian method to match people. It is fast but can swap IDs when people overlap. DeepSORT adds an appearance feature, so it can match people again after they are hidden, but it needs more computing. ByteTrack [3] also uses low-confidence detections, which helps when athletes partly hide each other. Trackers are commonly tested with the CLEAR MOT metrics. DeepSORT needs an extra model and has trouble with similar uniforms. ByteTrack depends on the quality of the detector and, in its basic form, has no appearance cue.

VI. HUMAN POSE ESTIMATION

Today, multi-person pose methods are of two types. Top-down methods find each person first and then estimate the pose inside the box. Examples are RMPE/AlphaPose and HRNet, which is known for accurate points. Bottom-up methods find all body points in the image and then group them into people. OpenPose [1] does this in real time using part affinity fields. A third group is made for small devices: BlazePose and MediaPipe [2] are light and fast, but the MediaPipe pose solution works for one main person, so for many athletes we need a separate detector and one crop per person. Pose methods are usually compared using PCK or average precision. The cost of OpenPose does not rise much with more people, but it can make grouping errors in crowds and usually needs a GPU. Top-down methods are accurate and handle imperfect boxes, but their cost rises with more people and they depend on the detector; HRNet is also a heavier model. Light methods such as BlazePose and MediaPipe run on a CPU or mobile device, but may be less accurate in extreme poses or when the person is hidden.

VII. MOVEMENT ANALYSIS IN SPORT

After the body points are found, technique can be checked in two ways. Rule-based analysis compares joint angles, distances or timing with ranges set for an action. It is clear and easy for a coach to review and change, but experts must set the limits for every sport and action. Learning-based analysis uses models such as ST-GCN to recognise actions or quality. It can learn complex patterns, but it needs labelled data and is harder to explain.

VIII. SPORTS SYSTEMS AND DATASETS

Thomas et al. [4] review computer vision in sports. They name player tracking, pose analysis and event detection as main uses, and blocked views, camera motion and similar uniforms as ongoing problems. AI Coach [5] uses pose estimation to help personal athletic training. Most existing sports datasets and systems use broadcast cameras and one sport, so they do not directly match low-cost, fixed-camera training sessions. AI Coach [5] is the closest to our goal; Athletix AI adds multi-athlete tracking and coach queries.

IX. OPEN PROBLEMS

- Blocked views and ID swaps: athletes overlap, cross each other and wear similar uniforms, so tracks can break.
- Linking a track to a real athlete: a track ID is not a real identity. Names set by the coach, jersey-number reading or other consent-based methods are needed.
- Sport-specific meaning: a general pose model cannot decide what a mistake is. Rules differ by sport and action and need expert review.
- Limits of 2D view: angles from one camera depend on camera position and are less reliable for movements toward or away from the camera.
- Real-time limits: accuracy must be balanced with speed and hardware cost, especially with many athletes.
- Fast motion and light: motion blur, low light and far-away athletes reduce quality.
- Fairness: results may differ by body type, clothes, age and ability, so testing must include different groups.
- Privacy and consent: recording athletes, including minors, needs consent, limited access and limited storage time, for example under India's Digital Personal Data Protection Act, 2023.
- Reliable evidence: many systems show results only in controlled conditions. A pose difference is not proof of a mistake, so feedback should show uncertainty.

X. FUTURE DIRECTIONS

- Sport-specific datasets recorded in real training conditions, with coach-labelled mistakes.
- Better tracking and re-identification for crowded groups in similar uniforms, with coach confirmation.
- Mixed methods that keep clear rules but use learned models for action recognition, tested for each sport.
- Feedback that shows confidence and lets the coach confirm, correct or reject it.
- Text and voice interfaces for coaches, based only on stored analysis data.
- Light models for small devices so that tools stay low-cost.
- Comparing the same athlete across many sessions.

XI. WHAT THIS MEANS FOR ATHLETIX AI

The review supports several design choices for Athletix AI, our proposed real-time multi-athlete coaching system. First, MediaPipe pose works for one person at a time, so a person detector and a tracker should give one crop per person to the pose model. Second, the tracker should be tested for ID swaps when athletes overlap, because individual reports need stable IDs. Third, tracks should be linked to athletes by coach-given names or jersey numbers, and the report should show uncertainty when tracking is lost. Fourth, coach-defined angle rules are a good start because they are clear and need no labelled training data; learned action recognition can be added later after

testing. Fifth, the system should be tested with the metrics seen in this survey (detection precision and recall, MOTA and ID switches, PCK, latency and FPS) and with coach ratings. The prototype is still being built, so we do not claim any performance results.

XII. CONCLUSION

This survey reviewed the techniques needed for camera-based coaching support: person detection, multi-person tracking, pose estimation and movement analysis, along with existing sports systems and datasets. Good methods exist for each step, but reliable tracking of each athlete, sport-specific checking and ethical use are still open problems. These findings guide the design and testing plan of Athletix AI and show where more research is needed.

ACKNOWLEDGMENT

The authors thank their project guide, Prof. Aboli More, and the Department of Artificial Intelligence and Machine Learning, Zeal Polytechnic, Pune, for their guidance and support.

REFERENCES

[1] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “OpenPose: Realtime multi-person 2D pose estimation using part affinity fields,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, pp. 172–186, 2021.

[2] C. Lugaresi et al., “MediaPipe: A framework for building perception pipelines,” arXiv:1906.08172, 2019.

[3] Y. Zhang et al., “ByteTrack: Multi-object tracking by associating every detection box,” in Proc. ECCV, 2022.

[4] G. Thomas, R. Gade, T. B. Moeslund, P. Carr, and A. Hilton, “Computer vision for sports: Current applications and research topics,” Comput. Vis. Image Underst., vol. 159, pp. 3–18, 2017.

[5] J. Wang, K. Qiu, H. Peng, J. Fu, and J. Zhu, “AI Coach: Deep human pose estimation and analysis for personalized athletic training assistance,” in Proc. ACM Multimedia, 2019.