Video Analytics for Volleyball: Preliminary Results and Future Prospects of the 5VREAL Project

Andrea Rosani, Ivan Donadello, Michele Calvanese*, Alessandro Torcinovich, Giuseppe Di Fatta, Marco Montali and Oswald Lanz

Libera Università di Bolzano, Piazza Università 1, Bozen-Bolzano, 39100, Italy

Abstract

This paper introduces a real-time action recognition and tactical-behavior mining system designed specifically for volleyball games. The system aims to provide data augmentation, video annotation and KPI extraction processes by accurately identifying various actions and action sequential patterns performed during volleyball matches. Leveraging advanced computer vision techniques, the system aims at automatically detecting and recognizing player actions and group actions in real time. Then, Process Mining techniques are used to extract tactical behaviors, in the form of temporal relations, among player actions. By providing precise annotations, the system significantly provides an instrument for volleyball game analytics and tactical analysis. This paper outlines the architecture and key components of the real-time action recognition and tactical-behavior mining system and presents some preliminary results on the performance of the proposed model.

Keywords

Video action recognition, data augmentation, video annotation, process mining, sports

1. Introduction

Over the past decade, action recognition in professional sport activities has rapidly gained popularity as a tool for a variety of tasks such as player performance analytics, computer-aided game refereeing, and the like. In response to this interest, several action recognition systems have been devised in the context of several sports, such as football, basket, rugby, etc.

In this context, this paper presents an action recognition system for volleyball game analysis. The preliminary results obtained during the activity focus on the detection of actions, events, and tactical behaviors in volley with the final objective of providing a reliable Ai-powered data augmentation system that can be used for the TV broadcasting of volley games in a real time scenario, as well as for off-line analytics activities, starting from the video collected by a multi view source and shared using 5G transmission.

The document is structured into several sections that outline in detail the study process and the developments obtained. First, a review of some particularly relevant works in the specific field is proposed. Then, methods and algorithms are described, along with some results of preliminary experiments on a public dataset [7].

1.1. Context: the 5VREAL Project

This paper describes the preliminary results obtained during the activity related to the project 5VREAL – *5G Volley Reality Experience & Analytics Live*, focused on the study and implementation of a system for the acquisition, analysis and transmission of video and analytics in the context of volleyball games and training sessions. The project aims to create a scalable solution, which can be used at all levels of competition, professional and amateur.

Two use cases are developed:

- *Fun Engagement*: This use case aims to *use* artificial intelligence algorithms to enrich the spectator's experience while watching the match with augmented reality information displayed in real time on the broadcasted videos.

Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI, May 29-30, 2024, Naples, Italy

* M. Calvanese contributed with work done during his Master Thesis project at UPC Barcelona with Prof. Carlos Andujar Gran.

0009-0008-2622-6776 (A. Rosani); 000-0002-0701-5729 (I. Donadello); 0009-0005-4103-0147 (M. Calvanese); 0000-0001-8110-1791 (A. Torcinovich); 0000-0003-3096-2844 (Di Fatta), 0000-0002-8021-3430 (M. Montali), 0000-0003-4793-4276 (O. Lanz)

- *Coach*: Use of the game & 'rhythm' for technical staff. After the game, the technical staff or directly the coach receives indications on positions, speed, trajectories, time intervals between touches and higher-level semantic information about the tactical behaviors of the team that can favor a more in-depth technical and tactical analysis.

The involvement in the project of industrial partners operating in the media production sector will enable a real application scenario to test the performances of the proposed solution. The project is funded by the Italian Ministry of Enterprises and Made in Italy, MIMIT under the *MIMIT FSC 2014-2020: Tecnologie 5G. Progetti di sperimentazione e ricerca – Piano di Sviluppo e Coesione 2014-2020.*

2. State of art in action recognition and tactical behavior for volley

The task of action/pose estimation involves analyzing video content to track one or more persons of interest and identify their key anatomical features, typically defined as keypoints [14], [26]. When multiple actors interact, the task is usually referred as Group Activity Recognition (GAR) [18], [19], [22].

GAR algorithms differ in how they model spatial and temporal information in videos. Some dated approaches apply recurrent models: [7] develops a hierarchical model based on two long-short term memory (LSTM) models, [13] proposes a recurrent neural network (RNN) model with attention mechanisms and semantic graphs, [3] generates a map of candidate regions of interest and uses an RNN architecture for temporal processing, and [24] adopts a top-down approach using Gated Recurrent Unit.

Other works focus on convolutional mechanisms: [2] develops a convolutional relational machine for GAR, [19] works on individual poses using one-dimensional convolutional neural networks.

Newer models like graph-based networks and Transformers are also employed: [25] uses a graph-based model for spatio-temporal relationships, designs a descriptor for crowded scenarios, and [10] [12] proposes a Transformer-based solution for processing spatial and temporal information.

To recognize tactical behaviors, techniques like sequence mining algorithms and Inductive Logic Programming are used ([21], [19], [23]). Works in this field include [9] and [11] for predicting complex events from football matches using Answer Set Programming and Subgraph Discovery. In our work, temporal pattern mining algorithms based on Linear Temporal Logics will be used, offering a different approach compared to the mentioned works.

3. Methodology and algorithms

3.1. General architecture of the system

The AI block consists of a set of algorithms required to a) identify the position and trajectory of the ball, b) identify the position of individual players, and c) detect and identify actions performed within a specific timeframe.

The acquisition of images for AI occurs through three iPhone 14 Pro devices mounted tripods with calibrated cameras, connected to a backend via 5G, producing synchronized SRT (Secure Reliable Transport) compressed video streams.

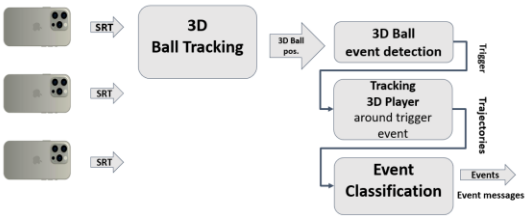

Figure 1: Overview of the architecture of the volleyball action recognition system.

The ball localization module starts the processing by producing a continuous data stream of the ball trajectory. When a change in its direction is detected, the player tracking and action detection modules are activated (Figure 1). This generates an output of the events occurred in the selected timeframe. In the following, we analyze in detail the different steps. 3D Ball tracking is described by a project partner in another submission to Ital-IA 2024.

3.2. Ball trajectories change detection

The general scheme for ball trajectory analysis can be subdivided in the following steps (Figure 2):

1. Identification of possible candidate ball positions.
2. Incremental interpolation of candidates with parabolic trajectories, producing a parabola for each frame.
3. Linking of trajectories from which to derive the motion of the ball.
4. Detection of trigger events when the ball undergoes an upward acceleration, such as a player touching or a bounce on the floor.

The algorithm, originally proposed in [5], requires as input the positions of the ball at each time step, that can be easily devised with a ball tracking system [14]. The path of the ball is modelled by a piecewise parabolic trajectory. Initially, seed triplets are identified within a threshold distance (r).

These triplets serve as initial anchors for parabolic fitting. Due to false positives, multiple seed triplets per frame may exist. Each triplet is used to fit a parabola, and candidate detections close to the

estimated position are added to a set of supporting points.

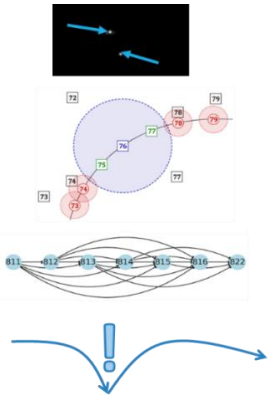

Figure 2: Ball trajectories analysis and trigger event detection [5].

The temporally furthest points within the support set are used to fit a new parabola. This iterative process continues until the set of supporting points ceases to grow. Parabolas with upward-pointing acceleration vectors are excluded as they violate physical constraints.

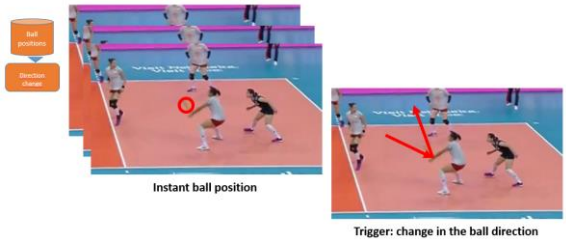

Figure 3: Action and Group Activity Recognition (images from [7]). The variation in the ball trajectory identifies an interaction that triggers the event.

To ensure a unique parabola per frame, trajectory distances are computed and used to construct a weighted graph. Dijkstra's algorithm [6] identifies the optimal path through this graph, yielding the final sequence of parabolas describing the ball's path.

Considering that the action mainly occurs around the ball's position, the proposed solution allows for detecting changes in the direction of the ball due to gameplay interactions. This trajectory variation triggers an analysis mechanism of the activities performed near the contact point to activate the subsequent phase of recognizing the actions of individual players and teams (Figure 3).

3.3. Individual player action recognition

In the rapidly evolving field of action recognition, many datasets, structures, and architectures have been introduced to address the challenges and complexities associated with understanding human actions in different environments [4]. These studies focus on extracting meaningful information from videos, by detecting and recognizing what a subject is doing [15], [16], [17].

The posture detection occurs within the video stream, in the player's bounding box, that is the area of interests of an object (the player, in this case) tracked in each video frame. The detection of the posture uses pose estimation technologies based on machine learning models [24], that identify key anatomical features of players, such as joints, extremities, center of mass, etc., commonly referred to as *keypoints* [8]. In the case of a volleyball player, the bounding box is used to locate the player's position within the video frame and subsequently extract keypoints on the players' bodies (Figures 4 and 5).

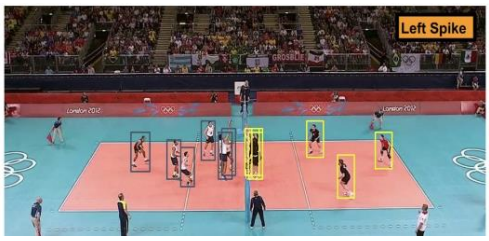

Figure 4: Example annotation from the Volleyball dataset showing the bounding box of each player divided by team (using different colors) and the action performed ("Left spike"). (Image from [7])

Starting from this information is possible to perform action recognition, as demonstrated effectively in [16], [17] that will be used as reference in the project for this specific task.

3.4. Team activity recognition

The challenge of Group Activity Recognition (GAR) requires addressing two main aspects. First, it demands a compositional understanding of the scene. Due to the relatively high number of people present in the scene, it's challenging to learn meaningful representations for GAR over the entire area. Since group activities often involve subgroups of actors and scene objects, the final label of the action depends on a compositional understanding of these entities. Secondly, GAR benefits from relational reasoning on scene elements to understand the relative importance of entities and their interactions [26].

4. Preliminary results

In the following, we present some preliminary results obtained using state-of-the-art techniques on public available datasets.

4.1. Dataset

The Volleyball dataset [7], represents a significant resource in the context of sports action recognition, specifically on volleyball. Although originally designed for athlete action recognition, the dataset

has been extended to include the task of 2D ball detection in the image. The dataset comprises a total of 4830 frames from 55 videos, offering a wide variety of actions and activities to analyze (Figure 4). In the dataset, there are nine annotations for individual player actions and eight group activities, detailed in Table 1.

Table 1

Classes of individual player activities are listed, and group actions, including the number of instances.

Action Classes	No. of Instances	Group Activity Class	No. of Instances
Waiting	3601	Right set	644
Setting	1332	Right spike	623
Digging	2333	Right pass	801
Falling	1241	Right winpoint	295
Spiking	1216	Left winpoint	367
Blocking	2458	Left pass	826
Jumping	341	Left spike	642
Moving	5121	Left set	633
Standing	38696		

4.2. Group activity recognition

GAR is performed at different levels. Initially, the keypoints of the various players are extracted. Based on these, an estimation of the action each player is doing is defined, and then related to the predicted level of person-to-person and person-to-group interaction.

4.2.1. Trigger event identification and GAR

The situation that activates the GAR mechanism is represented by the *trigger*, identified with the change of the ball direction (Figure 5).

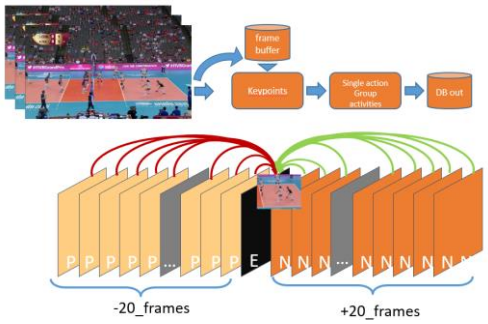

Figure 5: -Detailed schema for action and group activity recognition.

In Figure 6 we present some frames from [7], processed using the proposed algorithms, detailed in the following section, allowing for a comprehensive visualization of the keypoints of the various players combined with the trajectories of the ball

4.2.2. Hierarchy of semantic events for GAR

Taking inspiration from the approach proposed in [26], composite learning of entities in the video and relational reasoning on these entities is established.

Figure 6: Example 2D application of player identification and identification of ball trajectory changes ("trigger"). Keypoints can be observed on each player's silhouette, along with the corresponding arc of the ball trajectory.

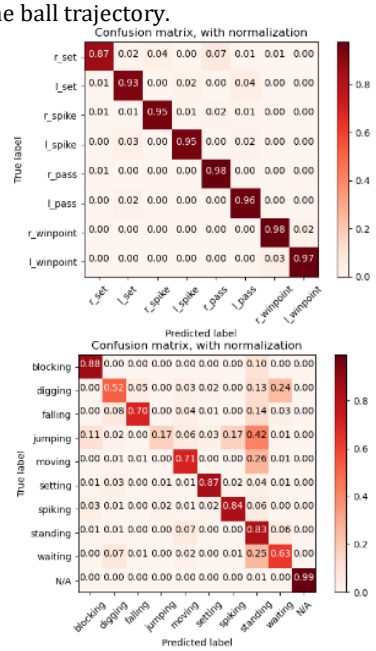

Figure 7: Our results on the Volleyball dataset considering the Olympic Split [7], [26]. In the first confusion matrix we represent GAR, in the second one the single player activities.

Like humans, object representation is performed at various granularities, as well as reasoning about their interactions to transform sensory signals into high-level knowledge. GAR is addressed by modeling a video as a set of tokens representing multi-scale semantic concepts present in the video, thus allowing the described method to be easily adaptable to understand any video with multi-actor multi-object interactions.

In the specific case of volleyball, the actors are represented by the players, while the object is represented by the ball. These tokens include keypoints, people, person-to-person interactions, person-to-group interactions, and object interactions. The performance of this analysis, compared to previous techniques based on standard RGB analysis

(i.e., considering the entire images and not just the keypoints), shows significant accuracy (Figure 7)

4.3. Tactical behavior

By tactical behavior, we mean a set of temporal relationships among volleyball actions that can lead to an outcome of particular interest, such as scoring a point. In what follows we provide a conceptual framework to formally define tactical behaviors and use Process Mining (PM) techniques for mining tactical behaviors from annotated volleyball matches.

4.3.1. A conceptual model for tactical behaviors

A tactical behavior is a set of temporal relationships over events in a volleyball match. An event is the main action of a player on the ball which has a start time, an end time, a set of players involved with information related to their pose, their bounding boxes, their unique identifiers, the quality of the action and the position of the ball. For example:

- A dunk by a player from area A1 is immediately followed by a point scored.
- A reception (with low quality) of a player is immediately followed by a point.

Our conceptual model for a volleyball event is shown in Figure 8.

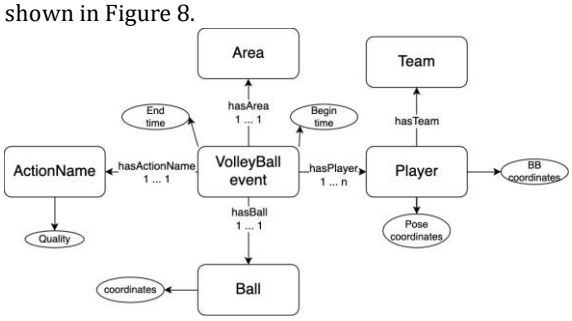

Figure 8: The conceptual model for volleyball events.

A volleyball match is therefore a sequence of annotations of volleyball events in chronological order. Such events are annotated with the use of the computer vision techniques above or provided by scoutmen.

4.3.2. Process Mining for tactical behaviors

Process Mining [20] embraces Data Mining and Knowledge Representation and focuses on the analysis and improvement of business processes based on data collected from the information systems. One of its key features is the availability of tools for mining information from temporal discrete data. We analyzed the matches of the Volleyball dataset (converted in a suitable format) with the Process Mining RuM (Rule Mining Made Simple) tool [1] to mine tactical behaviors.

RuM extracts temporal relations among actions of volleyball events through a list of templates defined with Linear Temporal Logic over finite traces (LTLf), one of the reference logics in the field [28]. Examples of such templates are the *Chain Response* between actions A and B that means that action A must be immediately followed by action B or the *Alternate Precedence* between A and B that means that action B must be preceded by action A without any other occurrence of B in between, see [27] Table 2. In addition, RuM provides the selection of a numeric support that indicates the percentage of occurrence of a particular template in the set of matches that can be used as a key process indicator. The 55 Volleyball matches were analyzed in less than 10 seconds, a suitable performance for an offline scenario. With a support of 20%, we obtained 50 tactical behaviors expressed using LTLf templates, automatically translated by the tool in natural language sentences for a better human comprehension. An example of mined tactical behavior is that in the 47.73% of the matches, each jump (for a block) is preceded by a dunk without any other jump in between. In addition, RuM also allows us to link the tactical behaviors of actions to the other concepts of the above conceptual scheme.

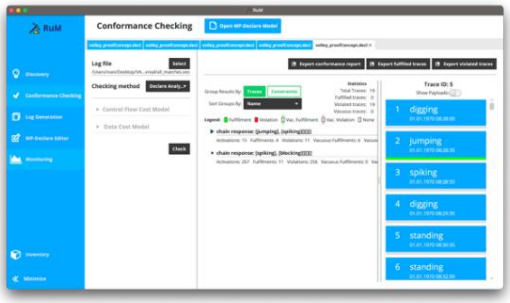

Figure 9: The conformance checking analysis of predefined tactical behaviors.

RuM also supports the manual definition of tactical behaviors and the analysis of the matches according to such predefined behaviors. This task is called *conformance checking* and, as two examples of tactical behaviors, we defined that a jump is followed by a spike and that a spike is followed by a block. Figure 9 shows the results of the RuM conformance checking.

Each behavior is analyzed for each match and, on the right, the actions of match 5 are shown and highlighted in green if they conform to the tactical behavior, in red otherwise.

Acknowledgements

This work is supported by 5VREAL – 5G VOLLEY REALITY EXPERIENCE & ANALYTICS LIVE, CUP I53C23001340005, funded by Italian Ministry of Enterprises and Made in Italy.

References

[1] Alman, A., Donadello, I., Maggi, F. M., Montali, M. Declarative Process Mining for Software Processes: The RuM Toolkit and the Declare4Py Python Library. In Int. Conf. on Product-Focused Sw Process Improvement (2023).

[2] Azar S.M., Atigh M.G., Nickabadi A., Alahi A.: Convolutional relational machine for group activity recognition. In: IEEE CVPR. (2019)

[3] Bagautdinov, T., Alahi, A., Fleuret, F., Fua, P., Savarese, S.: Social scene understanding: End-to-end multi-person action localization and collective activity recognition. IEEE CVPR. (2017)

[4] Camarena F, Gonzalez-Mendoza M, Chang L, Cuevas-Ascencio R: An Overview of the Vision-Based Human Action Recognition Field, Math. Comput. Appl. 2023

[5] Calvanese M: Ball tracking in Padel Videos using Convolutional Neural Networks. [Laurea magistrale], Università di Bologna, Corso di Studio in Artificial intelligence, 2023

[6] Dijkstra E.W: A note on two problems in connexion with graphs. Numerische mathematik, 1959

[7] Ibrahim MS, Muralidharan S, Deng Z, Vahdat A, Mori G. A hierarchical deep temporal model for group activity recognition. CVPR, 2016

[8] Jiang T, Lu P, Zhang L, Ma N, Han R, Lyu C, Li Y, Chen K: RTMPose: Real-Time Multi-Person Pose Estimation based on MMPose. ArXiv, 2023

[9] Khan, A., Bozzato, L., Serafini, L., Lazzerini, B. (2019). Visual reasoning on complex events in soccer videos using answer set programming. In GCAI 2019.

[10] Li J, Wang C, Zhu H, Mao Y, Fang H, Lu C.: CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark, CVPR, 2019

[11] Meerhoff, L. A., Goes, F. R., De Leeuw, A. W., Knobbe, A. (2020). Exploring successful team tactics in soccer tracking data. In Machine Learning and Knowledge Discovery in Databases: Int. Workshops of ECML PKDD 2019.

[12] Nabi, M., Bue, A., Murino, V.: Temporal poselets for collective activity detection and recognition. In: IEEE CVPR. pp. 500–507 (2013)

[13] Qi, M., Qin, J., Li, A., Wang, Y., Luo, J., Van Gool, L.: stagnet: An attentive semantic rnn for group activity recognition. In: Proc. of the ECCV. (2018)

[14] Rahimian P, Toka L: Optical tracking in team sports: A survey on player and ball tracking methods in soccer and other team sports. Journal of Quantitative Analysis in Sports, 2022

[15] Sudhakaran S, Escalera S, Lanz O: Gate-Shift Networks for Video Action Recognition, IEEE CVPR 2020

[16] Sudhakaran S, Escalera S, Lanz O: Gate-Shift-Fuse for Video Action Recognition, IEEE TPAMI, 2023

[17] Takahashi M, Ikeya K, Kano M, Ookubo H, Mishina T: Robust Volleyball Tracking System Using Multi-View Cameras. ICPR, 2016

[18] Thilakarathne H., Nibali A., He Z., Morgan S.: Pose is all you need: The pose only group activity recognition system (pogars). arXiv preprint arXiv:2108.04186 (2021)

[19] Van Haaren, J., Ben Shitrit, H., Davis, J., Fua, P. (2016, August). Analyzing volleyball match data from the 2014 world championships using machine learning techniques. In Proceedings of the 22nd ACM SIGKDD (pp. 627-634).

[20] Van Der Aalst, W., van der Aalst, W. (2016). Data science in action (pp. 3-23). Springer Berlin Heidelberg.

[21] Wenninger, S., Link, D., Lames, M. (2019). Data mining in elite beach volleyball–detecting tactical patterns using market basket analysis. IJCSS, 18(2), 1-19.

[22] Wu L.F., Wang Q., Jian M., Qiao Y., Zhao, B.X.: A comprehensive review of group activity recognition in videos. International Journal of Automation and Computing pp. 1–17 (2021)

[23] Xia, H., Tracy, R., Zhao, Y., Fraisse, E., Wang, Y. F., Petzold, L. (2022, November). VREN: volleyball rally dataset with expression notation language. In 2022 IEEE ICKG (pp. 337-346).

[24] Xu D., Fu H., Wu L., Jian M., Wang D., Liu X.: Group activity recognition by using effective multiple modality relation representation with temporal-spatial attention. IEEE Access 8, (2020)

[25] Yan R., Xie L., Tang J., Shu X., Tian Q.: Higcin: hierarchical graph-based cross inference network for group activity recognition. IEEE TPAMI (2020)

[26] Zhou H, Kadav A, Shamsian A, Geng S, Lai F, Zhao L, Liu T, Kapadia M, Graf HP: COMPOSER: Compositional Reasoning of Group Activity in Videos with Keypoint-Only Modality. ECCV, 2022

[27] Donadello, I., Di Francescomarino, C., Maggi, F. M., Ricci, F., Shikhizada, A. Outcome-oriented prescriptive process monitoring based on temporal logic patterns. Engineering Applications of Artificial Intelligence (2023).

[28] Claudio Di Ciccio, Marco Montali: Declarative Process Specifications: Reasoning, Discovery, Monitoring. Process Mining Handbook 2022.