An Offline Neuro-Symbolic Football Pattern Retrieval Approach Using Constraint Programming

Augustin Crespin
ICTEAM, UCLouvain, Belgium

Pierre Schaus
ICTEAM, UCLouvain, Belgium

1 Abstract

The paper with the same title has been accepted for publication in the Main Track of CP 2026.

When preparing for upcoming matches, football coaches must create a well-defined game plan, typically requiring hours of video analysis of opponents' previous matches. This work is tedious, as analysts often rely on high-level aggregate statistics rather than targeted tactical search. Being able to search for specific tactics in a video feed, rather than re-watching entire clips, would significantly improve efficiency. Deep learning methods alone performs poorly at retrieving intervals in which specific spatio-temporal or tactical patterns occur. One might be tempted to use multimodal large language models for this, but preliminary experiments with VideoLLaMA 3-7B [4] revealed their limitations: when given a 30-second clip containing six passes and prompted to count them, the model gave incorrect and inconsistent answers (10, 14, at least 5, and even hallucinated a commentator stating 40 passes). This highlights the need for a dedicated system.

These observations motivated the development of TactiCP, a loosely-coupled neuro-symbolic approach, illustrated in Figure 1, in which precomputed frame-level detections are processed through an SQL-like domain-specific query language. The pipeline relies on three distinct processing units:

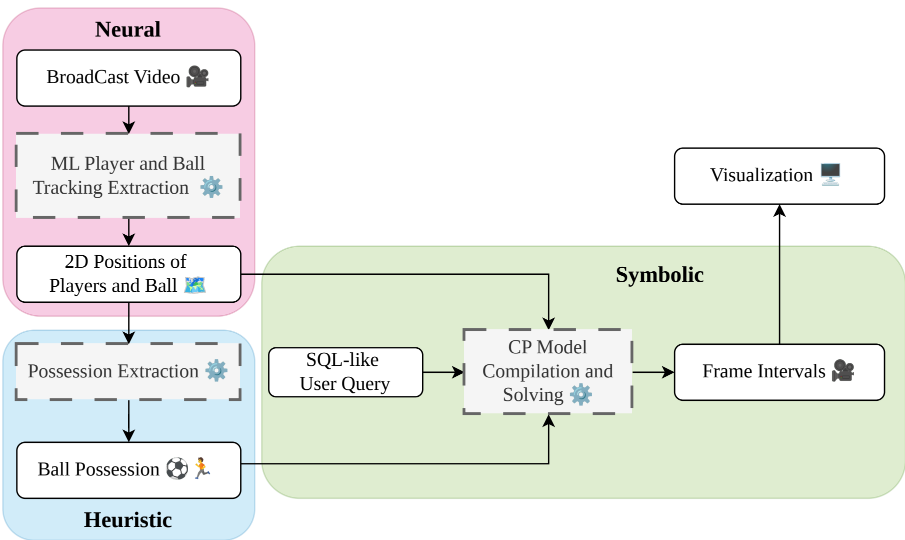

Figure 1 Dataflow: Dashed rectangular nodes represent processing units (neural, heuristic, and symbolic), while rounded nodes denote input/output data artifacts.

42nd Conference on Very Important Topics (CVIT 2016).
Editors: John Q. Open and Joan R. Access; Article No. 23; pp. 23:1–23:3
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1. **ML Tracking (Neural):** The raw broadcast video is fed frame-by-frame into an ML tracking model, which extracts the continuous 2D positions of the players and the ball.
2. **Possession Extraction (Heuristic):** The raw positional data is passed to a heuristic module that determines which player, if any, possesses the ball at each individual frame.
3. **CP Compilation & Solving (Symbolic):** The user expresses a tactical pattern query via a Domain-Specific Language (SQL-like DSL syntax). Based on this query and the discrete data (positions and possession), a CP model is compiled and solved to retrieve the frame interval(s) satisfying the query constraints.

The method leverages well-established CP constructs, such as interval variables and regular constraints adapted for interval variables. This paper focuses exclusively on the heuristic and symbolic components of the pipeline, relying on ground-truth annotations from the SoccerNet Game State Reconstruction task [3] for the first stage.

To intuitively understand the system, consider a user who wants to find, in a game video, all sub-video intervals where Player 11 passes the ball to Player 16, with each interval starting exactly when the ball leaves the foot of Player 11 and ending when it arrives at Player 16. The user could formulate the following simple query:

```
SELECT(PLAYER("player11",11).PASSTO(PLAYER("player16",16)).MINRANGE())
```

Under the hood, TactiCP translates this high-level query into a Constraint Satisfaction Problem (CSP), whose solutions correspond to one or more intervals of video frames satisfying the query constraints.

The contributions of this work are threefold: (1) an SQL-like domain-specific language for spatio-temporal tactics; (2) a compilation technique into a CP model relying on the known interval variables [1] and the regular constraint [2] adapted for interval variables; and (3) an open-source tool and experimental evaluation, accessible at `tacticp.com`.

Experimental evaluations on real football games demonstrate that the system is both expressive and computationally viable. Pass detection performance was compared against the winner of the 2023 SoccerNet Ball Action Spotting Challenge [1] [2]. Both models correctly identified four out of five passes; however, TactiCP produced no false positives while the ML model generated four. On this example clip, TactiCP achieves a precision of 1.0, a recall of 0.8, and an F1 score of 0.89, compared to precision 0.5, recall 0.8, and F1 score 0.62 for the ML model. This illustrates the typical behavior of the two approaches: the ML model produces more detections but also more false positives, while TactiCP generates fewer but more precise detections. Computational efficiency is equally notable: across 164 clips, the ML model required on average 7.82 seconds per clip, while TactiCP required only 0.095 seconds, roughly two orders of magnitude faster.

Future work will explore the integration of frequent pattern mining algorithms, so that automatically mined patterns could be formalized, queried, and analyzed in depth using an extended version of the proposed DSL.

[1] The rules and development kit for this challenge can be found **here**
[2] The source code can be found **here**

References

1. Philippe Laborie and Jerome Rogerie. Reasoning with conditional time-intervals. In *Proceedings of the 21th International Florida Artificial Intelligence Research Society Conference, FLAIRS-21*, pages 555–560, 01 2008. URL: http://www.aaai.org/Library/FLAIRS/2008/flairs08-126.php.

2. Gilles Pesant. A regular language membership constraint for finite sequences of variables. In Mark Wallace, editor, *Principles and Practice of Constraint Programming – CP 2004*, pages 482–495, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg. doi:10.1007/978-3-540-30201-8_36.

3. Vladimir Somers and et al. SoccerNet game state reconstruction: End-to-end athlete tracking and identification on a minimap. In *2024 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, jun 2024. doi:10.1109/CVPRW63382.2024.00334.

4. Boqiang Zhang, Kehan Li, Zesen Cheng, Zhiqiang Hu, Yuqian Yuan, Guanzheng Chen, Sicong Leng, Yuming Jiang, Hang Zhang, Xin Li, Peng Jin, Wenqi Zhang, Fan Wang, Lidong Bing, and Deli Zhao. Videollama 3: Frontier multimodal foundation models for image and video understanding. *arXiv preprint arXiv:2501.13106*, 2025. URL: https://arxiv.org/abs/2501.13106, doi:10.48550/arXiv.2501.13106.