Machine Learning for Time Series

Presentation - Mini-Project

Adonis JAMAL   Fotios KAPOTOS

SoccerCPD: Formation and Role Change-Point Detection in Soccer Matches Using Spatiotemporal Tracking Data

École Normale Supérieure Paris-Saclay - MVA

December 17th 2025

MATHÉMATIQUES
VISION
APPRENTISSAGE

Context: The Fluidity of Soccer

The Challenge in Sports Analytics

- **Dynamic Formations:** Teams constantly reshape (e.g., 4-4-2 → 3-5-2) based on game state.
- **Noisy Data:** Players swap positions temporarily (overlap, pressing).
- **The Gap:** Standard methods often assume static formations or are too sensitive to noise.

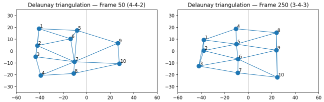

Figure: Formation shifts captured via Delaunay triangulation.

Project Goal

- Reproduce and extend **SoccerCPD** [1], a change-point detection framework.

Work Repartition & Novel Contributions

1. **Reproduction & Implementation**
   - **Full Pipeline:** Python package with R backend (for `g-segmentation`).
   - **FormCPD:** Delaunay-based spatial change detection.
   - **RoleCPD:** Permutation-based role switch detection.
2. **Novel Extensions (Beyond the Paper)**
   - **Possession Context:** Separating Attacking vs. Defensive shapes.
   - **Player Stationarity Metric:** Quantifying role fluidity (Fixed vs. Roaming).
   - **Event Data Generalization:** Adapting the framework for sparse StatsBomb data.

FormCPD: From Trajectories to Graphs

1. Role Assignment

- Players assigned to latent roles (Gaussian Mixture Models).
- One-to-one mapping via Hungarian Algorithm.

2. Delaunay Triangulation

- Encodes spatial topology $A(t)$.
- Permutation invariant.
- Distance: $d_M = \|A(t) - A(t')\|_{1,1}$.

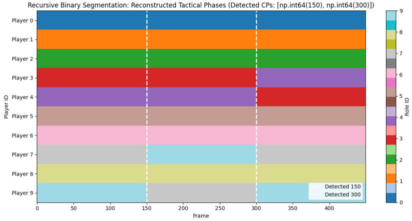

Figure: Topological changes trigger detection.

RoleCPD: Detecting Intraphase Swaps

Objective: Detect permanent role swaps (e.g., Wingers switching sides) within a stable formation.

The Metric: Switch Rate

- Input: Sequence of role permutations π_t.
- **Hamming Distance:**

$$d(\pi_t, \pi_{t'}) = \frac{1}{N} \sum_p \mathbb{I}_{\pi_t(X_p) \neq \pi_{t'}(X_p)}$$

- Measures deviation from the "Dominant Permutation".

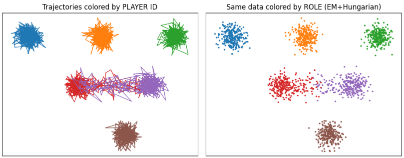

Figure: Synthetic Role Swap Data

RoleCPD: Recursive Segmentation

Discrete g-segmentation on Permutations

1. **Pre-process:** Filter frames with Switch Rate > 0.7 (temporary noise).
2. **Recursive Search:**
    - Apply graph-based change-point detection on the sequence.
    - **Stopping Criteria:** p-value > 0.01 or segment length < 5 min.
3. **Result:** Identifies distinct tactical role configurations.

High-Frequency vs. Sparse Event Data

1. Tracking Data (GPS)

- **Source:** Fitogether (10Hz) [1], SkillCorner (10Hz) [2].
- **Quality:** Dense, clean signals.
- **Prep:** Normalized to centered metric pitch.

2. Event Data (StatsBomb)

- **Source:** On-ball events only.
- **Challenge:** Extreme sparsity ($<$ 10 pts/min vs 600+ for GPS).
- **Bias:** Spatial bias toward the ball.

Novel Adaptation for Event Data:

- **Pseudo-Trajectories:** Aggregated events over 5-minute rolling windows.
- **Proxy Graph:** Average adjacency matrices instead of frame-by-frame Delaunay.

Density Comparison

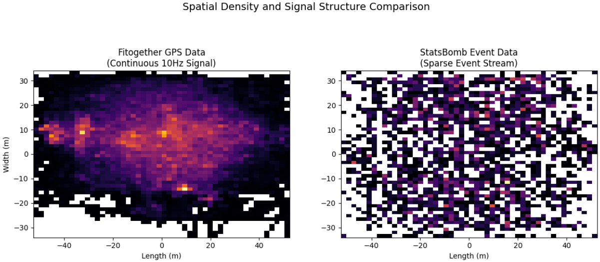

Figure: Heatmap: Dense GPS coverage vs. Sparse/Biased Event Data.

Robustness: Sensitivity Analysis

Parameter: `min_fdist`

- Controls threshold for declaring a new formation.
- **Finding:** Higher threshold acts as a pruning mechanism.
- **Result:** Major tactical shifts remain stable; minor noise is filtered out.

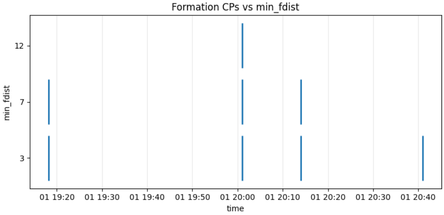

Figure: Timeline stability across thresholds.

Contextualizing Formations (GPS)

Problem: A single "average" centroid formation ignores game state.

Method:

- Filter frames by possession status.
- Compute centroids separately.

Insight:

- Defense: Compact block.
- Attack: Wide, expansive shape.

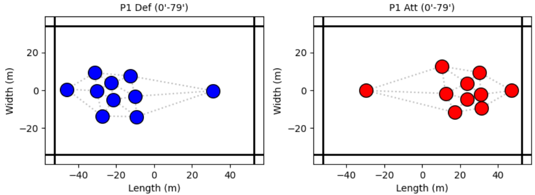

Figure: Match 1925299: Distinct shapes detected.

Quantifying Role Fluidity (σ)

Definition: Standard deviation of Euclidean distance from the role center.

GPS Data Findings (A-League)

- **Fixed Roles:** CBs show low σ (Stationary).
- **Roaming Roles:** Wingers/Midfielders show high σ.
- Validates tactical roles beyond simple coordinates.

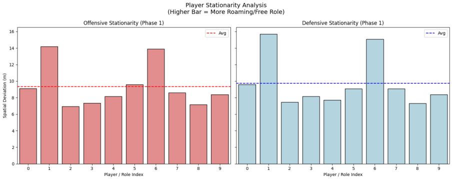

Figure: Stationarity profile (GPS Data).

Offensive vs. Defensive Structures (FC Barcelona)

We extended the "Possession Context" analysis to sparse StatsBomb data.

Tactical Shapes in Sparse Data

- Despite low data density, distinct shapes emerge when aggregated over time.
- **Defense:** High density in central zones (low width).
- **Attack:** Wingers push high and wide; full-backs overlap.

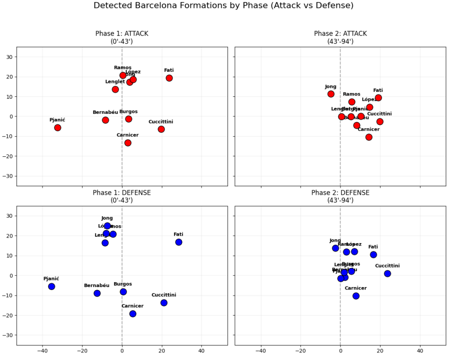

Figure: Barcelona: Attack (Red) vs Defense (Blue).

Quantifying Fluidity in Event Data

Does the "Stationarity Metric" hold up with sparse event streams?

Comparison Results

- **Consistency:** The "Fixed" vs. "Fluid" distinction persists.
- **Phase Analysis (Barcelona):**
  - *Attack Phase:* High deviation (Avg 12m) → Fluidity.
  - *Defense Phase:* Lower deviation (Avg 9.6m) → Rigidity.

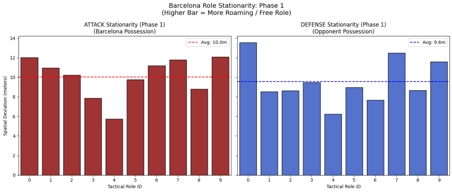

Figure: Stationarity: FC Barcelona (Event Data).

Conclusion

- **Successful Reproduction:** Implemented FormCPD and RoleCPD, validating the framework on synthetic and real matches.
- **Robustness:** Confirmed method stability via sensitivity analysis.
- **Key Extensions:**
  1. **Possession Context:** Revealed distinct attacking/defensive shapes.
  2. **Stationarity Metric:** Provided quantitative insights into player role adherence.
- **Versatility:** Demonstrated applicability to sparse event-stream data.

Bibliography I

[1] Hyunsung Kim et al. "SoccerCPD: Formation and Role Change-Point Detection in Soccer Matches Using Spatiotemporal Tracking Data". In: *The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, 2022, pp. 3146–3156. DOI: 10.1145/3534678.3539150. URL: https://doi.org/10.1145/3534678.3539150.

[2] SkillCorner. *SkillCorner Open Data: Broadcast tracking data.* https://github.com/SkillCorner/opendata. Accessed: 2025-01-04. 2020.

RoleCPD: Hamming Distance Switch Rate

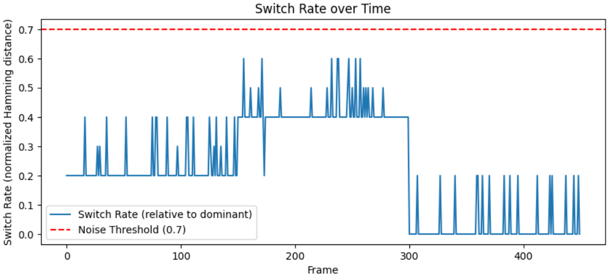

Figure: Hamming distance-based switch rate time series used for change point detection.

RoleCPD: Formation Ground Truth

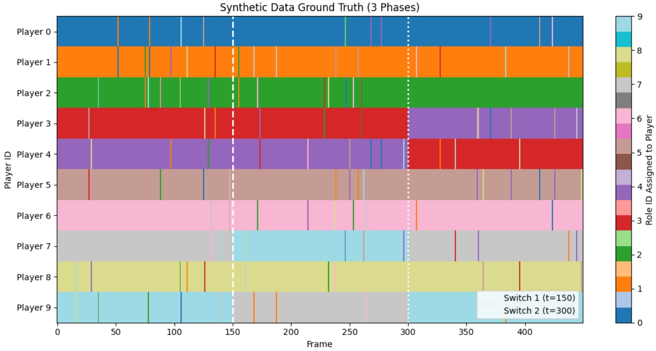

Figure: Ground truth role assignments over time with two tactical switches at frames 150 and 300.

RoleCPD: Need for Recursive Segmentation

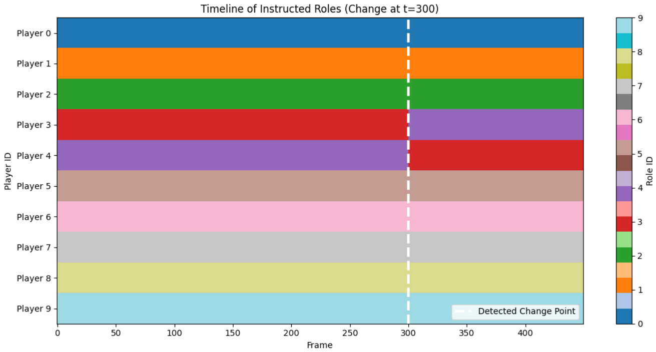

Figure: Visualization of one detected tactical phase with dominant role assignments. This shows the need for recursive segmentation to capture multiple phases.

Stationarity: FC Barcelona Phase 2 (Event Data)

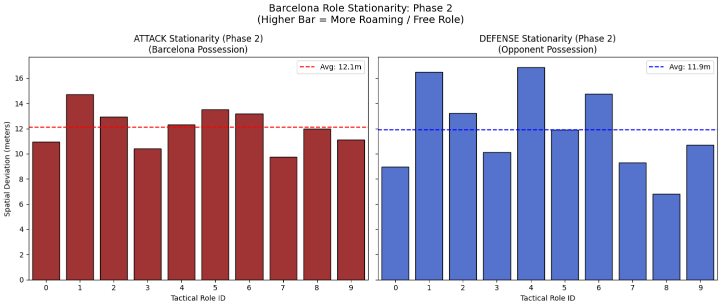

Figure: Stationarity: FC Barcelona (Event Data).