electronics

MDPI

Article

Comparing Tactical Analysis Methods in Women's Soccer Using Positioning Data from Electronic Performance and Tracking Systems

Luis Ángel Oliveira *, David Melendi and Roberto García

Department of Informatics, University of Oviedo, 33203 Gijón, Spain; melendi@uniovi.es (D.M.); garciaroberto@uniovi.es (R.G.)
* Correspondence: oliveiraluis@uniovi.es; Tel.: +34-619817454

Abstract: Although, in recent years, it has been common to monitor players in team sports using EPTSs (Electronic Performance and Tracking Systems) devices, most of the studies have focused on the optimization of individual performance rather than collective work or tactical analysis. Moreover, almost all these studies focus on men's teams with little focus on women's teams. In this work, data from women's soccer teams at different levels (competition and grassroots) have been collected using both a low-cost personally developed EPTS and a commercial EPTS. With these systems, we have built a dataset consisting of more than 16 million records, paying special attention to spatio-temporal variables collected in the form of geographical coordinates. Different methods have been applied to the collected dataset to solve the problem of determining the position (individual role) of each player on the field based solely on spatio-temporal variables. The methods include algorithms based on clustering, centroid calculation, and computer vision. We have verified the effectiveness of these methods and propose an alternative method based on image recognition algorithms applied to heat maps generated from the position of the players monitored during the matches. As shown in this paper, the validity of the proposed method has been verified, exceeding the performance of existing methods and extending the range of application of these techniques.

Keywords: EPTS; grassroots football; centroids; heatmap; women's soccer; k-means

check for updates

Citation: Oliveira, L.Á.; Melendi, D.; García, R. Comparing Tactical Analysis Methods in Women's Soccer Using Positioning Data from Electronic Performance and Tracking Systems. *Electronics* **2024**, *13*, 1876. https://doi.org/10.3390/electronics13101876

Academic Editor: Young-Joo Suh

Received: 10 April 2024
Revised: 8 May 2024
Accepted: 8 May 2024
Published: 10 May 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Historically, the performance of players in team sports has been the subject of multiple studies, although most of them have focused more on individual analysis than on the collective. Generally, they seek to evaluate individual performance by assessing individual variables (physical or emotional), such as strength, speed, technical skills, fatigue, etc. [1–3]. Moreover, from a gender perspective and considering soccer practice, almost all these studies have focused on men's teams, with very few references to women's football and grassroots football in general, and there are few references in this regard, with all of them being produced very recently [4]. The traditional approach to tactical study was based on observation during competitive matches by expert personnel (scouters), with their evaluation being subject to subjective knowledge. Therefore, this qualitative evaluation is generally subjective, person-dependent, and unsystematic, with the subsequent analysis being carried out manually, which could take considerable time. However, in recent years, due to the increasing use of EPTSs [5–7], the explosion of big data, and growing computing capacity, it is possible to conduct an objective and systematic analysis based on the collected data in near-real time. This, together with other factors, such as the growing interest in the tactical analysis of modern football, has led to an increase in studies aimed at obtaining metrics and indices that allow us to predict the behavior of players based on their positioning. It must be considered that all EPTSs incorporate some kind of global positioning system, which provides valuable information about the movements of the

players and their distribution on the pitch at any given moment, both in training and in competition. Many of their actions present collective behavior patterns, probably because of the work carried out during training sessions and based on rules set by the coaching staff, and as shown in some studies, it is possible that many of these patterns are predictable [8]. But again, we face the problem that many of these EPTS systems are prohibitively expensive, which means that these studies often only consider elite professional soccer, and most of the time only apply to men's football, despite the demonstrated growth of women's football in recent years [9]. Although there are different ways to determine individual roles and predict where each player will move in the next moment, this work will only consider the spatio-temporal variables of each player, ignoring others such as the position of the ball or the positions of the opponents. This would not be feasible in the context to which this work is applied due to the lack of financial resources (in the case of the ball) or others (the positions of the opponents).

Therefore, the first goal of this study is to validate some of the metrics proposed for elite men's professional football and to check if these metrics are also applicable to women's football and women's grassroots football when determining the tactical positioning of the team on the pitch. A new method obtained solely from the analysis of the players' spatio-temporal positioning data will be proposed to determine the team's playing scheme with the aim of improving collective performance to correct future situations and avoid the occurrence of undesired events. This method, unlike others, offers a new way of determining the tactical scheme of the team from images (heat maps), allowing the calculation of this scheme through the application of computer vision algorithms. Moreover, it will provide improvements to other methods already used in the past. For its application and testing, we have worked on data collected from under-represented groups such as women's football teams and grassroots football teams.

This article is structured as follows: Initially, the problem to be solved will be defined (Section 2). Subsequently, the methods generally used to solve the problem will be analyzed, applying several of them to the dataset developed in this work (Section 3). We will also explain how the data have been collected, the electronic devices used for this task, and the population sample used in constructing the dataset (Section 4). Some of the different proposed methods will be applied, and their results will be discussed in Section 5. An alternative method will be proposed as a solution to the original problem and its validity based on the constructed dataset. After that, in Section 6, a comparison between applied methods will be introduced. Finally, the conclusions and future improvements to this work will be presented (Section 7).

2. Problem Definition

In team sports, there are two approaches when carrying out a tactical study: an individual approach, based on the movements of each player according to his position (role), and a collective approach that takes into account other more general behaviors such as interline or intraline relations. Therefore, the first problem to be solved would be the determination of the individual role of each player within the team based on spatio-temporal data. Considering Bialkowski's definition in [10], the player's role or position on the field is understood as a space or area that is the responsibility of each player within the team. These roles, in many sports, such as basketball or football, generally have names assigned to them (center back, midfielder, striker, etc.), although sometimes there is a certain disparity in the face of small strategic, geographical, or cultural variations for the same position (for example, in this context, midfielder, interior, or winger could be considered synonyms). From the individual roles, it would be possible to derive the tactical disposition of the team, for example, if a 1-4-3-3 or 1-4-2-3-1 formation is used. Taking into account only the spatio-temporal variables, we can say that the position of each player at a given instant in time will be given by his spatial coordinates: $x_t = (x_i, y_i)$. Thus, knowing the successive coordinates of the player throughout a match or training session would allow us to easily represent her trajectory and, therefore, her movements, as shown in Figure 1.

This figure shows the sampling points of a player during a training exercise. Similarly, the set of positions occupied by all team members at a given time t will be represented by a set of tuples of the form $x_t = \{(x_1, y_1), (x_2, y_2), \ldots, (x_{11}, y_{11})\}_t$, where the subscript uniquely identifies each player.

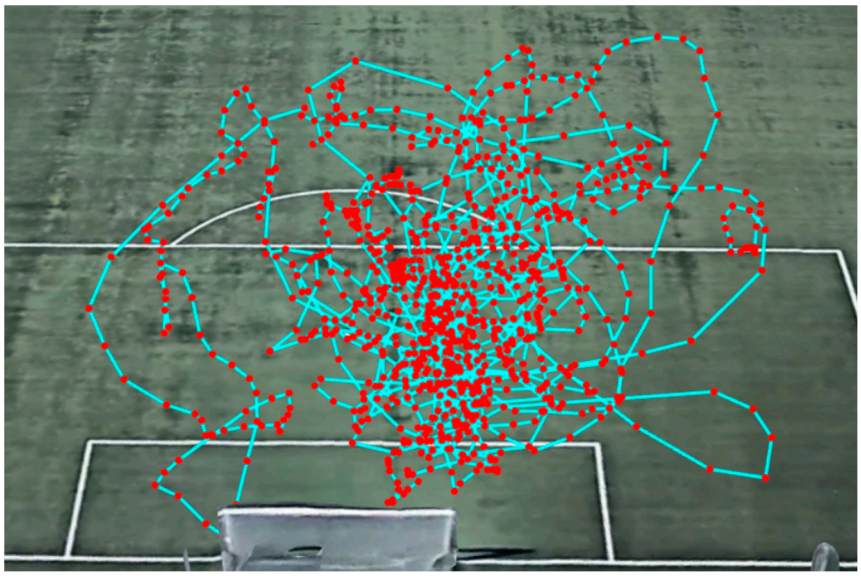


Figure 1. Trajectories of a player with her point cloud during a training session.

One approach to solving this initial problem of determining the individual role of each team member could be based on the heat map of each player, as shown in Figure 2. This heat map represents the densities of the points occupied by the player during a period.

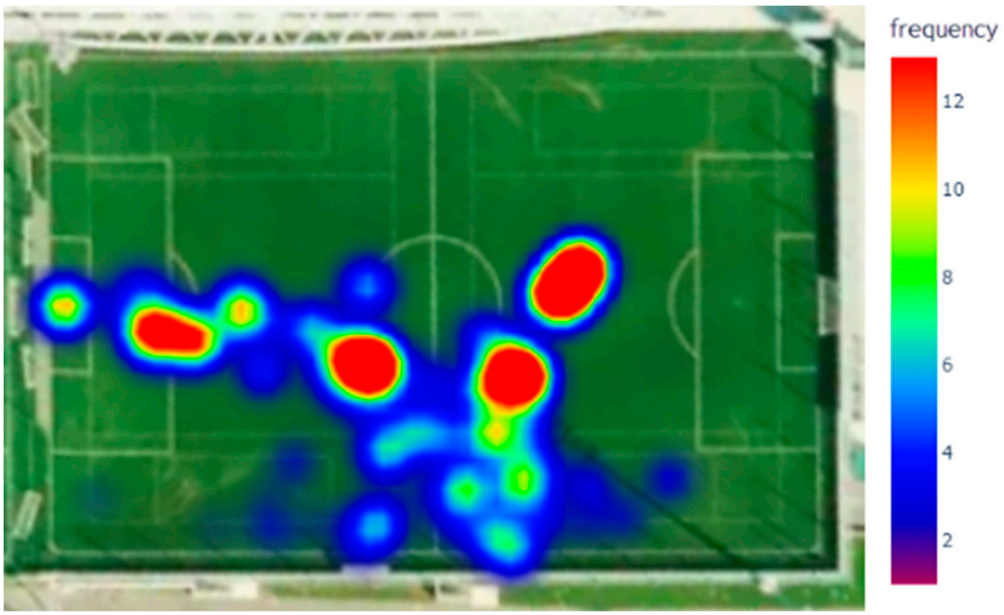


Figure 2. Heat map of a female player during a match.

In some cases, based solely on this information, the role of each player within the team could be determined exactly. However, as observed in the previous figure (Figure 2), sometimes it is impossible to determine this role precisely due to different situations. For example, a player may switch sides during a match, occupy different positions in specific actions (corners in favor or against, direct free kicks, etc.), or enjoy freedom of

movement during a match. This is easily noticeable in Figure 3, which shows a segment of the trajectories occupied by a player during a training match (in this figure, colors represent the speed range).

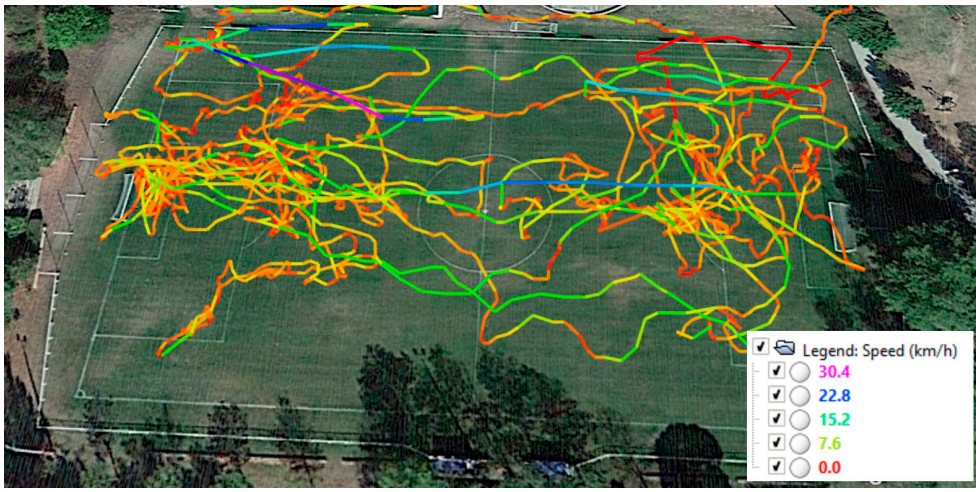


Figure 3. Player trajectory during a training match. Colors represent different speeds.

All this makes it less easy to determine the individual role based on her heat map. In fact, Figure 2 corresponds to a player who occupies the position of right back, and yet, the heat map does not allow us to extract such information. Therefore, this first solution would not be valid, and other possibilities would have to be explored. The next problem to solve, once the individual roles have been identified, would be to see if there are any dependencies when predicting the future movements of the players. Sometimes, there are dependencies marked by these roles. For example, it is common for the defensive line to move following a reference player who acts as a guide.

In other instances, it is the strategies set by the coach that dictate these dependencies. For example, during an attacking phase on the right flank, the left back should remain in line with the center backs in anticipation of a possible loss of possession, or during a set-piece, certain players must provide cover against a potential counterattack. It is also essential to consider that the movements and positions of the players may vary depending on whether it is an attacking or defensive phase.

Finally, once the potential metrics to set dependencies between players are established and validated, these results might be utilized for three objectives: to (a) predict future player movements, (b) correlate them with critical events (goals, chances against, etc.), and (c) identify and rectify tactical aspects to improve collective tactical behavior.

3. Related Work

In their article Large-Scale Analysis of Soccer Matches Using Spatio-Temporal Tracking Data, Bialkowski proposes a method that enhances role assignment based on spatio-temporal variables, known as minimum entropy data partitioning [10], wherein the role of each player is defined in relation to the roles assigned to the rest of the teammates for a specific moment. This approach defines a formation as a spatial arrangement of players from a strategic viewpoint, which can be characterized as a set of individual roles. Each role within a formation is unique (i.e., no two players in a team can have the same role simultaneously), but players can exchange roles during the match.

Additionally, there can be multiple formations during the same match, which can be interpreted as different sets of roles at different moments. The goal is to determine the role of each player using clustering techniques. Like the previous case, another clustering-based method is proposed in [11]. In this case, the aim is to find the role of each player using a two-phase intelligent algorithm: in the first phase, the optimal number of clusters is calculated. In the second phase, the centroid of each cluster is computed to determine the

role of each player. Other methods ([12,13]) propose solutions based on the calculation of the centroids by partitioning the surface and using the average of the positions occupied over time by each player or by analyzing a specific situation.

In Refs. [14,15], three new approaches for tactical analysis are proposed: measurement of the coordination between players (measuring the interline and intraline distances of the field players in successive instants, as shown in Figure 4), team coordination in critical events, and interaction with the opposing team, although the latter is not possible to address in the present work due to the lack of such information.



Figure 4. Situation of the players at a given moment. Women's grassroots football team. Data collected through self-developed ETPS.

In Ref. [16], Voronoi diagrams are employed to identify and investigate the spatial dynamics of player behavior in futsal. In Ref. [17], such positional data are used to uncover structures and certain patterns in the game, such as corners, direct free kicks, etc., also by employing clustering techniques. In other studies [18], space-time information is classified using decision trees to obtain the tactical arrangement of players. As evident, tactical analysis based on spatio-temporal variables is a field of increasing interest, using which clubs and teams seek to gain an advantage over their opponents. This paper offers another approach to this problem: obtaining the tactical scheme from images (heat maps) by applying computer vision algorithms.

4. Materials and Methods

For the preparation of this article, data were collected from different women's teams, both in training matches and in competitive matches. The teams monitored ranged from grassroots football teams (alevin or U-12, infantil or U-14 and regional categories) to the second category of the Spanish Royal Spanish Football Federation (2nd RFEF), which is already considered semi-professional. The population used in data collection, as far as women's football is concerned, was organized according to the age ranges shown in Table 1.

Two different EPTSs were used for data collection. EPTSs are widely used for individual performance monitoring and analysis in different team sports [5] and have been the subject of numerous studies that have demonstrated their effectiveness [19]. Since a commercial system is usually expensive, and small clubs do not usually have EPTS available daily, a self-developed EPTS was used for data collection at the grassroots level, which proved to be as efficient as a commercial system [20]. The aim was to build an inexpensive and simple device that would allow teams without resources to monitor their players, obtaining results like those of the professional EPTSs available on the market. The minimum requirements that were determined for the device were as follows: (1) portable

and small in size, (2) equipped with GPS/Wi-Fi, (3) support for data storage (typically micro-SD card), (4) very low cost and (5) autonomous (battery). On the other hand, it was intended to fit into the DIY (do-it-yourself) philosophy, for which it had to be very easy to build.

Table 1. Sample population.

Category	Number of Players	Age (Mean)	Age (SD)
2nd RFEF	20	19.05	4.53
Nacional	19	16.16	3.10
Regional	26	14.77	0.83
Infantil	32	12.84	0.60
Alevín	30	11.20	0.67

During the process of selecting the device to be built, different options were proposed. For testing purposes, we initially worked with a prototype consisting of a device made up of an Arduino nano, a GPS NEO 6M, a micro-SD card, and a power supply. Although it would be desirable to have a GPS module that could operate at 10 Hz, such as the NEO M9N, as it is considered optimal for geolocation, the NEO 6M was chosen for cost reasons. Other open hardware-based devices, like Raspberry or Maduino, were also tested. For power supply, different approaches were considered, from CR3022 button cells to power bank or Lipo batteries. Tests were carried out in different outdoor sports (football, running, cycling), with the intention of verifying that the future device could fulfill the desired requirements. The device was tested in different wearable supports (bracelets, vests, etc.), with all of them being satisfactory. Initially, we worked on collecting information on an individual basis to later extrapolate the results to the group. After building different prototypes, the selected hardware was Maduino-based. Concretely, the microcontroller was an IoT Maduino Zero A9G. This little motherboard accomplished all requirements posed. Maduino Zero A9G is an IoT (Internet of Things) solution based on the 32-bit Atmel's SAMD21 MCU and GPRS/GSM GPS module A9G. It integrates a microcontroller ATSAMD21G18, GRRS/GSM GPS module A9G, power management, and storage. A Lipo 1800 mAh/3.7 V battery was added to the supply power, as it is capable of uninterrupted operation for more than 20 h. Although the data could be transmitted in real time using a SIM card, we chose to store it locally on each device, using a micro-SD card. Also, a case was manufactured via 3D printing to carry the set. The device shown in Figure 5a was used to collect data from grassroots soccer teams. For the professional category, data were collected using the OHCOACH Cell EPTS shown in Figure 5b [21].



(**a**) (**b**)

Figure 5. (**a**) Low-cost EPTS. (**b**) OHCOACH Cell EPTS.

A brief comparison of both devices is shown in Table 2.

Table 2. EPTS comparison.

	Low Cost EPTS	**OHCOACH EPTS**
Cost	USD 40	USD 200
Size	54 (L) × 80 (H) × 16 (T) mm	45 (L) × 76 (H) × 18 (T) mm
Weight	55 g	51 g
Battery	>20 h (GPS only) USB charge	6 h (GPS only)
Storage	Depends on micro-SD used	128 MB NAND
Global positioning	1 Hz GPS	10 Hz GNSS
Wireless communication	GPRS (download 85.6 Kbps, upload 42.8 Kbps)	Wi-Fi 802.11b/g/n and Bluetooth
Live Monitoring	Requires SIM	Requires OHCOACH Live Hub
Operation Condition	−40 °C to 85 °C	−10 °C to 55 °C

However, given that the main purpose of this work is tactical analysis, we preferred to primarily work with data collected during competition matches in the aforementioned 2nd RFEF category. Data from 28 matches for all participating players were used. The OHCOACH Cell device operates at a frequency of 10 Hz, providing a minimum of 594,000 location points per match for the 11 players. The complete dataset for a single team contains more than 16 million records of global positioning. Tactical data from lower categories teams were only considered for purely comparative purposes.

The data for each individual match were cleaned prior to processing, as there were stretches of time when the data were not useful for tactical analysis. Generally, players switch on their devices some time before the start of the match, so the positioning data includes warm-up and stoppage periods that could add noise and lead to modified or undesired results in the tactical analysis. Therefore, only the actual competition sections were considered. The data stored in each device were identified by a unique identifier (usually the bib number) and downloaded at the end of the session, whether derived from training or competition. Once the data were cleaned, they were analyzed using Python scripts, as this allows easy processing and graphical representation, following some of the existing metrics. The results and conclusions of each method are shown below.

5. Results Analysis

As we have seen in previous sections, there are different approaches applicable to obtain the role of each of the players. Some of them have been applied to the reference dataset. The Python scripts, an example dataset, and an example image used in the following sections for the implementation of the different methods have been published on the following website: https://github.com/loliveirar/EPTS, accessed on 7 May 2024. The results obtained are shown below.

5.1. Clustering-Based Methods

Different clustering methods have been applied to the initial dataset but have not been successful. Different methods have been tested: DBSCAN (Density-Based Spatial Clustering of Applications with Noise), GMM (Gaussian Mixture Model), and K-means. Of these, the one that has provided the best results is based on K-means. K-means is a popular clustering algorithm used to partition a dataset into a predefined number of clusters. The algorithm iteratively assigns each data point to the nearest cluster centroid and then recalculates the centroids based on the mean of the data points assigned to each cluster. This process continues until the centroids no longer change significantly or a specified number of iterations is reached. The following figure (Figure 6) shows the results after applying this method to the data of a competitive match (2nd RFEF category team).

Goalkeeper data are not considered because the goalkeeper never wears EPTS devices to prevent injuries. The arrow represents the direction of attack. The following process has been followed to obtain the graph shown: starting from the spatio-temporal data collected by the EPTS, the K-means class of the Python scikit-learn library has been used to obtain the scatter plot shown, using $n = 10$ as the number of clusters to obtain (corresponding to the number of players on the field) and a random seed of reproducibility *seed* of 42.

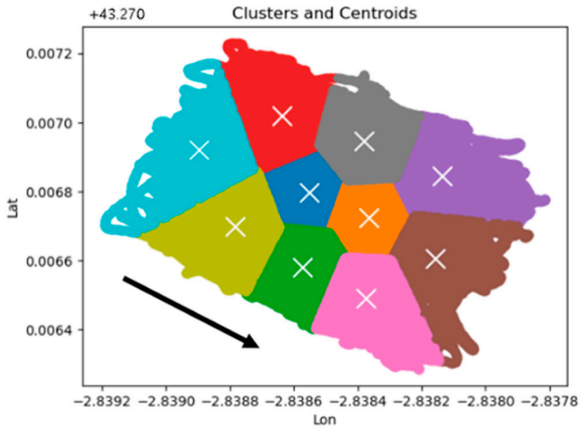


Figure 6. Clusters obtained applying K-means (2nd RFEF category team). Each color represents a cluster for each player. The symbol x is the centroid of each cluster. (Lezama, 2 December 2023).

The reproducibility seed is used to ensure that the results of a machine learning algorithm, such as K-means in this case, are reproducible. When running a machine learning algorithm that relies on random processes, such as random centroid initialization in the case of K-means, the results may vary between different runs, even if the same dataset and parameters are used.

By setting a reproducibility seed, the initial state of the random number generator used by the algorithm is fixed. This means that even though the processes within the algorithm depend on randomness, the result will be the same for each run when using the same seed.

Visually, it can already be seen that the calculated clusters differ from the 1-4-2-3-1 scheme used by the team for the sample match. Therefore, this method has not been effective in inferring the roles of each player from the dataset under study. When applied to a grassroots football team (U-14 category), the result, as expected, is even less accurate, as shown in Figure 7.

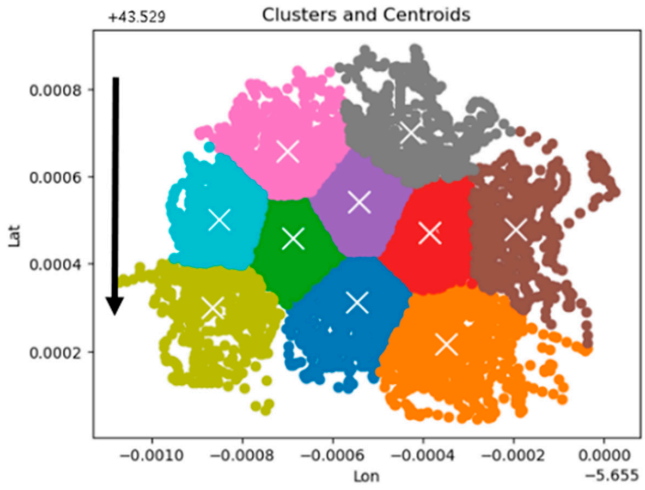


Figure 7. Clusters obtained by applying K-means (U-14 team). Each color represents a cluster for each player. The symbol x is the centroid of each cluster. (Gijón, 26 June 2023).

During this match, the monitored team played with a 1-4-3-3 formation. As mentioned above, tactical discipline is less in these categories; hence, the results are even worse.

5.2. Centroid-Based Methods

The centroids have been studied in numerous articles [22–25] as a tool with great potential for tactical analysis. The centroid of a set of geographical coordinates, as in the case at hand, identified by their longitude and latitude coordinates, will be the average point that represents the center of mass of the set. Mathematically, the centroid can be calculated by taking the average of the coordinates of all points in each dimension. To calculate the centroid of each player, starting from her individual geopositioning data, the average position of all the positions occupied during the match by that player has been calculated. Similarly, the centroid by lines and the total centroid of the team have been calculated. This approach provides a point that minimizes the sum of the squared distances to all other points in both dimensions. By calculating the centroid of each player throughout each match, said centroid will give us their mean position. Applying this method to all participating players in the match would yield the arrangement of each of them, thus allowing for the identification of each player's role within the team and, therefore, their tactical disposition. Figure 8 shows the centroids calculated for each player over four different matches. Each point represents the calculated centroid for each player. Colors represents different lines: defenders, midfielders, and forwards. The direction of the arrow indicates the direction in which the team attacks. Note that in the two top figures, the obtained points have been projected onto real maps. However, this is not possible in all situations due to map lack of updating or displacement errors generated by GPS. In these cases, a template image has been used for clarity purposes. Although the use of a template may cause some distortion (not all fields have the same dimensions), the initial positioning of the equipment has been taken as a reference to eliminate such distortion.

As can be seen in the displayed images, this method proves highly effective in obtaining the team's tactical disposition from spatio-temporal variables. It is observed that the team maintains an organized 1-4-2-3-1 formation and does not vary its playing scheme throughout the different matchdays of the regular league. It is worth noting that the goalkeeper never wears EPTS devices to prevent injuries. This analysis also serves to draw other conclusions, both individual and collective. From an individual standpoint, it could be observed if a player deviates excessively from their theoretical position or in relation to their teammates in the same line (for example, in Figure 8a, it is observed that the midfielders were too close together). From a collective perspective, other conclusions can be drawn: if the block has been positioned high (Figure 8b) or low, if the block has been compacted or not (open or closed), or the interline distance (for example, comparing Figure 8c,d, it is observed that in matchday 12, the block was higher than in matchday 13). From an intraline or interline perspective, the method of calculating centroids by lines, as shown in Figure 9, or even for the entire team, can also be applied.

When this method is applied to grassroots football teams, it is observed that the positions are still not so fixed, as shown in Figure 10. This is true in the different cases in which it has been calculated, in line with the findings of [13], which suggests that centroids vary with the age of the players (both at the individual and team level), implying a better tactical performance that improves with age.

In this case, it can easily be deduced that this is a tactical arrangement in a 1-4-3-3 formation. However, the structure is not as well defined, perhaps because the players have not yet established and internalised the concepts of spatial distribution on the pitch as clearly as professional teams.

In conclusion, it seems that this method provides a fairly adequate result in obtaining the tactical disposition of the team from spatial–temporal variables, being more reliable the higher the level of the players. It also provides a working and improvement tool for the coaching staff of teams in development.

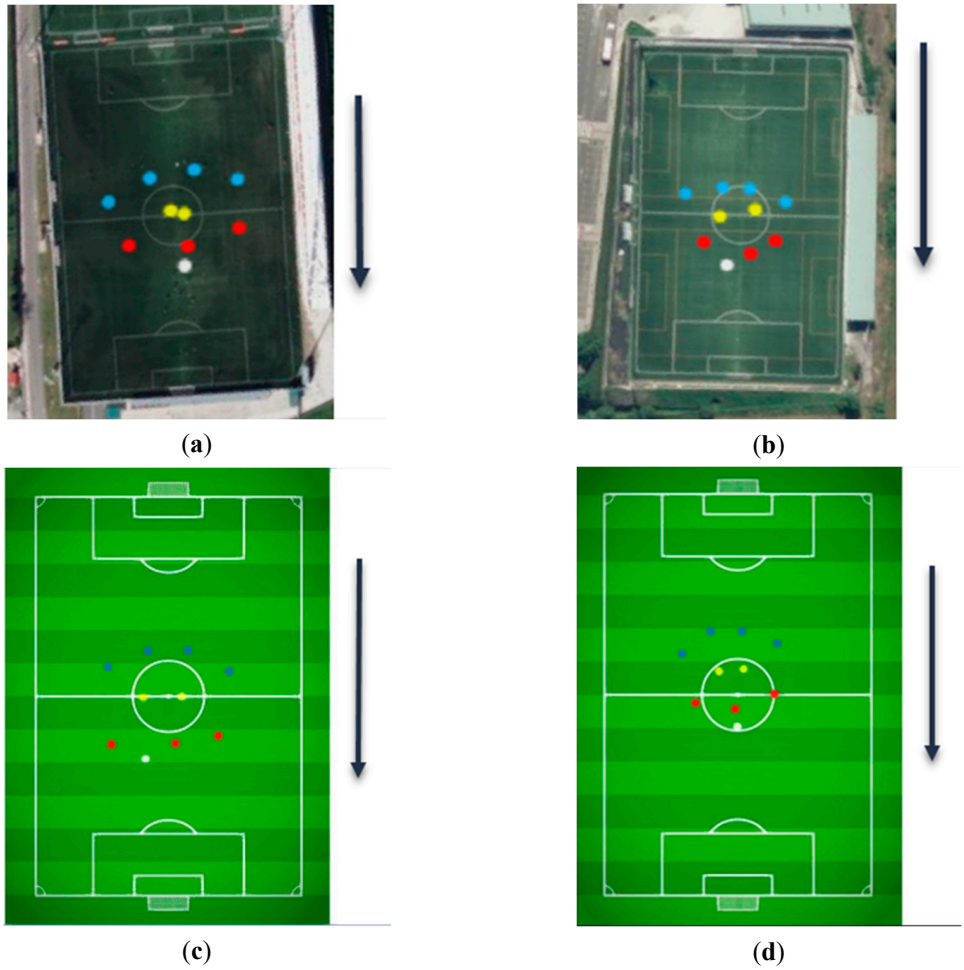


Figure 8. Centroids of each player over four matchdays of the regular league 2023/2024. Each color represents a different line within the team: defenders (blue), pivots (yellow), midfielders (red) and forwards (white). Arrow indicates the direction of attack (**a**) Matchday 10. (**b**) Matchday 11. (**c**) Matchday 12. (**d**) Matchday 13.

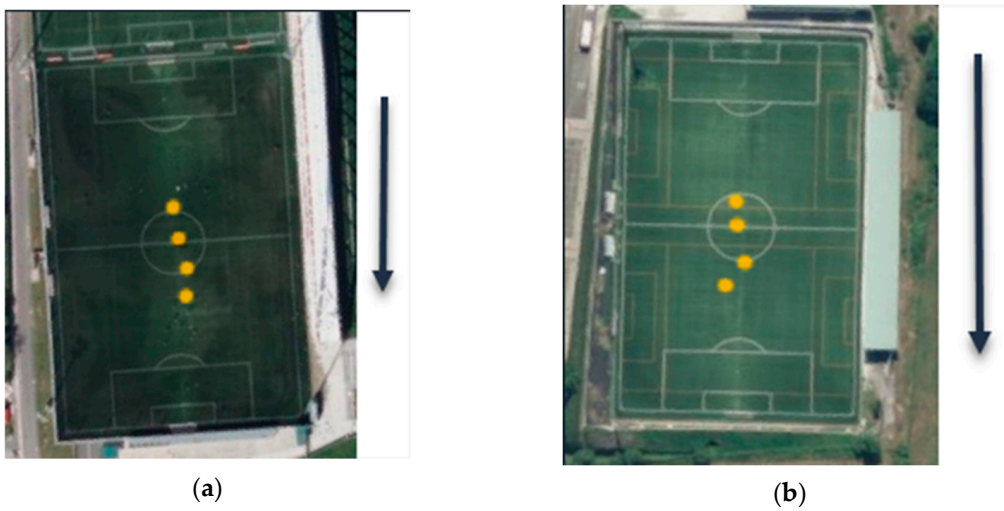


Figure 9. Intraline centroids (one per line). Arrow indicates the direction of attack. (**a**) Matchday 10. (**b**) Matchday 11.

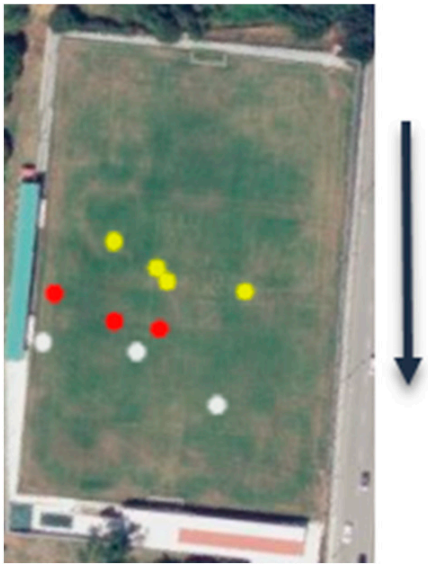


Figure 10. U-14 team centroids. Each color represents a different line within the team: defenders (yellow), midfielders (red) and forwards (white). Arrow indicates the direction of attack.

5.3. Method Based on Image Analysis

The method proposed below, unlike previous ones, is based on the use of computer vision algorithms to obtain the players' roles from an image. There are some studies based on image analysis applied to other sports aimed at obtaining occupation maps on the field [26]. Based on the succession of points occupied by a player throughout the match, it is possible to generate her heat map. However, on other occasions, this map can be obtained from other monitoring methods, such as the use of tracking cameras [27]. A heat map is a visual representation of data where color is used to indicate the intensity or density of certain values at specific points on the map. This type of visualization is useful for highlighting spatial patterns or trends. The information is presented using colors that vary in intensity or hue, usually on a color scale ranging from cooler shades (such as blue) for low values to warmer shades (such as red) for high values.

Each coordinate on the field has its corresponding representation, and the color intensity at each point reflects the magnitude of the value being represented, as shown in Figure 11. These heat maps allow for some parameterization, so that they can be adjusted to visually highlight the most frequented points by each player and analytically extract the top n points with the highest density. It could be argued that from these points (highlighted in red in Figure 11), the different roles of each player could be identified by comparison with the considered natural positions for each role. Depending on the team's scheme, these natural positions may vary, but they are well known for each scheme, as depicted in Figure 12.

This figure corresponds to the natural positions occupied by the players in a 1-4-2-3-1 formation, which, as previously mentioned, is the one commonly used by the team whose data are represented. In this figure (Figure 12), the right-back role is represented by the number 2. Through comparison with Figure 11d, the four main zones represented in red are shown. Taking n = 4 and calculating the mean position between these values, we observe that the obtained value deviates considerably from the theoretical position for the right-back role. In this specific case, it happens that this player is responsible for occupying the closing position in all set-piece plays that occur in attack. Hence, there are two hot zones in the opponent's half. The same applies to the rest of the players represented in Figure 12.

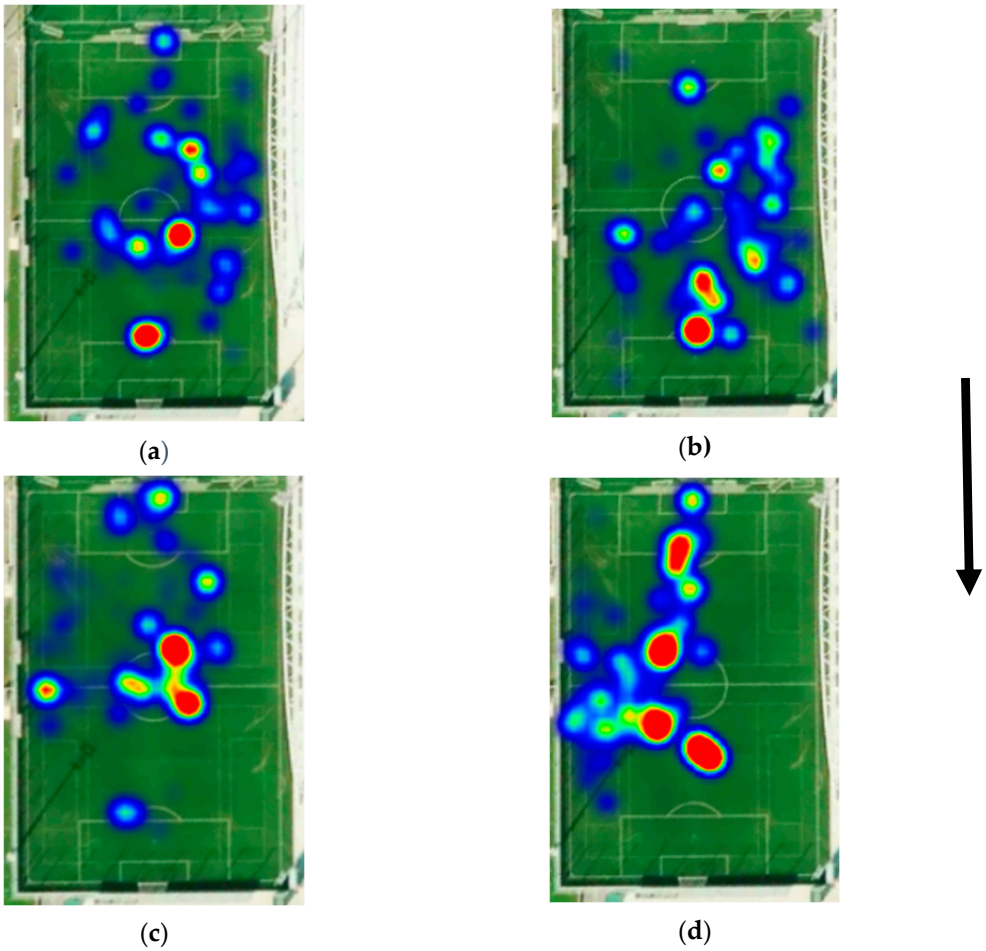


Figure 11. Heatmaps depending on the player's role. (**a**) Defensive midfielder. (**b**) Support striker. (**c**) Center back. (**d**) Right back.

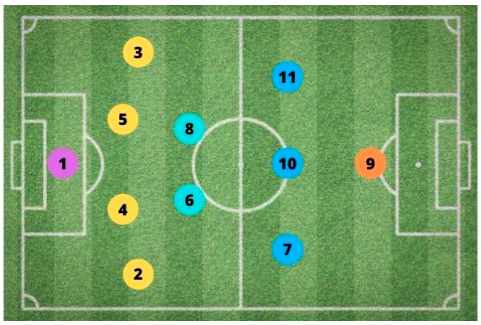


Figure 12. Typical players positions for a 1-4-2-3-1 formation. Each color indicates a line within the team.

Heat maps are a very useful tool when it comes to knowing the distribution of players on the field of play and can be applied at the line level (defensive, pivots, midfielders, and forwards) or at the team level, providing valuable information about field occupation and player distribution over time, as shown in Figure 13. This allows for assessing whether the team played more in one zone of the field or the other, whether they played with a high or low defensive line, and other tactical parameters (for example, in the represented case, it is observed that the team attacked more on the right flank than on the left flank).

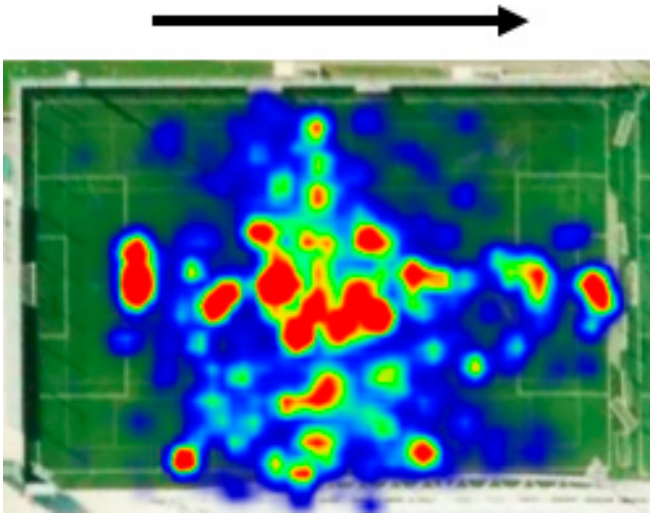
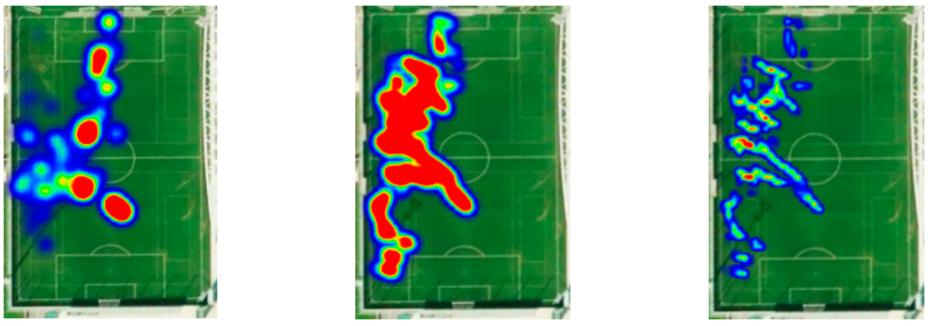


Figure 13. Team's heat map.

However, as a conclusion, it is deduced that although heat maps provide useful information about individual and collective performance, in the analyzed case with the given dataset, based on the information presented so far for this method, they have not proven to be an effective tool for calculating roles based on spatio-temporal variables. Therefore, the proposed improvement would be to try to eliminate those lapses in which the player remains static for reasons of stationary play (e.g., waiting for a corner kick). A player is considered to be "stationary" when there is less than 0.20 m between two consecutive coordinate measurements (Figure 14a reveals that situation). After eliminating these time sections (Figure 14b) and adjusting the radius of influence of the points in the heat map representation (Figure 14c) (from r = 12 to r = 5), the centroid of the hottest zones is now adjusted to the natural position for the role. In this case, it is the right back. When representing a heat map, the radius of influence of the points plays a crucial role in determining the accuracy and granularity of the heatmap representation. The radius of influence essentially defines the area around each point where its intensity affects the visualization of the heatmap. A larger radius of influence means that each point will have a broader impact on the surrounding area, resulting in smoother transitions between intensity levels and potentially obscuring smaller variations in the data. On the other hand, a smaller radius of influence will result in sharper transitions between intensity levels and highlight finer details in the data but may lead to a more granular representation that can be harder to interpret. In summary, the radius of influence of the points in a heatmap influences the level of detail and smoothness of the visualization. Adjusting this parameter is essential for striking the right balance between highlighting important patterns and avoiding the loss of valuable information in the data.



(**a**) (**b**) (**c**)

Figure 14. Proposed improvement based on heat maps. (**a**) Original heat map (r = 12). (**b**) Refined heat map (r = 12). (**c**) Final heat map (r = 5).

Calculating the centroid of the n hottest zones (n = 4), the point obtained corresponds more accurately to the natural position for the player's role, as shown in Figure 15. The values selected here for r, n, and the thresholds have been established after multiple experiments, and no major variations were observed when the thresholds or the number of hot (red) zones were modified, although the algorithm was designed to obtain as many relevant zones as possible. The radius of influence r is relevant from a visual point of view, but if the geographical coordinates are used as a dataset for the subsequent analysis, it would have no impact on the results obtained.

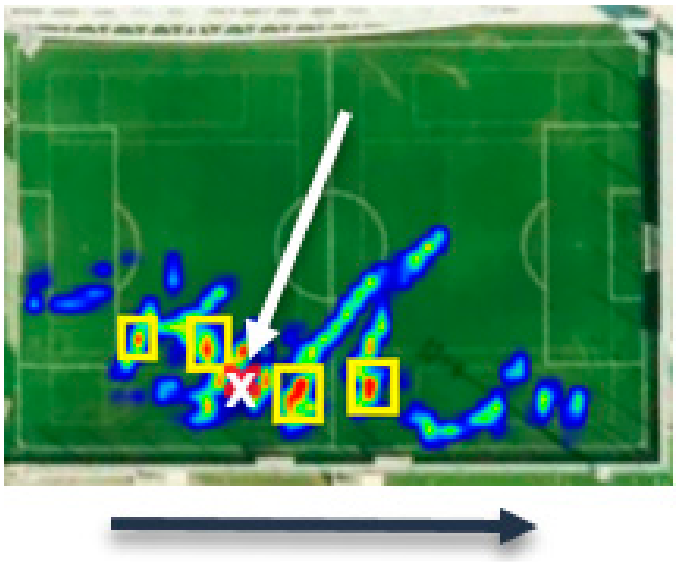


Figure 15. Result after refining the heat map. Hottest zones are within the yellow squares (n = 4) and the new centroid is marked in white.

The proposed method allows for working directly with heat maps, whether constructed from a dataset or not. This presents the advantage of faster computation for determining the individual role of each player (processing large data collections typically require a lot of time and resources) by generally working directly with images and a simplified derived set from the original dataset (if available). Therefore, another advantage is that it does not necessitate having the original dataset. Currently, using image perception algorithms, it is relatively straightforward to extract the n hot zones from the image and, from this set of n points, calculate the centroid for each player. By adjusting the number n of points, the color thresholds used to detect (in this case, the red color), and the parameter r (in the image being analyzed), optimal results are obtained for the problem at hand. In a way, this method can be considered a combination of partition-based or clustering methods, but in this case, it operates directly on the images. Table 3 below details the result of the application of the proposed method on the image shown in Figure 15. Note that, in this case, each point indicates a pixel in the image represented as (x, y) and not a geographical coordinate. The theoretical point for the position of the right back would be around (40,90) with some margin of error, as not all teams play with the full backs at the same height. As can be seen on the original heat map, the result is very distant from the real one (± 42, ± 30).

Table 3. Results from the original image.

r	n	Upper Red Color Threshold (RGB)	Lower Red Color Threshold (RGB)	Obtained Points	Calculated Centroid Point
12	4	(255,0,0)	(255,75,75)	(100,172) (74,108) (70,155) (84,46)	(82,120)

However, when applied to the already optimised image shown in Figure 14c, the results are quite close to reality (±10, ±8), as shown in Table 4. In either case, the position is clearly identified, as it is the closest point to the centroid.

Table 4. Final results.

r	n	Upper Red Color Threshold (RGB)	Lower Red Color Threshold (RGB)	Obtained Points	Calculated Centroid Point
5	4	(255,0,0)	(255,75,75)	(60,55) (55,85) (42,104) (44,150)	(50,98)

In summary, this method to determine the tactical scheme of the team can be applied in two different ways, once the initial dataset has been refined: (a) On one hand, if the geographical coordinates of a player are available, the n areas with the highest density of points can be calculated, and then the centroid of these areas can be calculated. (b) On the other hand, if the geographical coordinates are not available, but a heat map is available, computer vision algorithms can be used to detect the n hottest areas and then calculate the centroid of these areas. In both cases, the result obtained improves on clustering-based methods and methods based solely on centroid calculation.

For this method to be effective, it is important to carry out a proper cleaning of the data to be represented. First, the positioning data to be considered should be restricted to match times. Likewise, 'stopped player' situations should be eliminated as closely as possible to avoid undesired effects on the result. Another aspect that must be considered when working with images is the translation of the geopositioning data to their corresponding coordinates in the image, since if this is not carried out properly, it could lead to erroneous results.

6. Comparing Methods

After carrying out comparative tests (eight matchdays) between the different methods to determine the tactical scheme of the team, the following conclusions have been drawn:

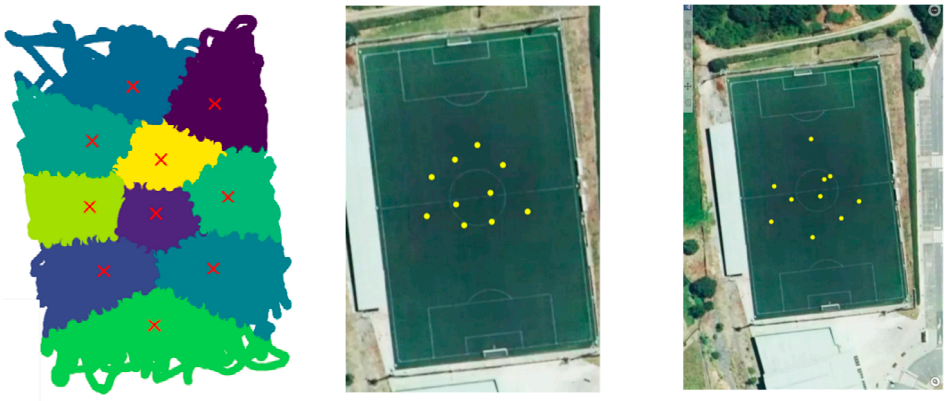
When working with heat maps, the proposed method is more than 90% faster in determining the centroid of each player, as the algorithm only has to extract the n hottest areas of the image, which is significantly faster than computing a dataset of more than 28,000 records per player/match (on average).

When working with a dataset, after the proposed data cleaning (removing non-useful data), the size of the dataset is reduced from 28,000 records per player/match to about 8700 records on average. After that, the n densest zones have been selected. It has been found here that the variations in n from four zones (values of n = 4, n = 10, and n = 20 have been tested) are not very relevant and hardly alter the results in most cases. In either case, the performance impact is significant as the size of the working dataset is considerably reduced.

The centroids obtained for each player are closer to the natural positions corresponding to the scheme used. These results are shown analytically in Table 5 (note that in the clustering-based method, the player/centroid correspondence need not be exact) and visually in Figure 16a–c. The deviation distances of each player from the theoretical position for the game scheme used (these theoretical reference positions are shown in Figure 12) during the first eight matchdays also have been calculated. The results of these calculations are shown in Table 6.

Table 5. Comparing methods. This table shows the calculated centroid (expressed by coordinates) using each method for matchday 12.

Player	Clustering Method		Centroid Method		Proposed Method (n = 20)	
	Lat.	Lon.	Lat.	Lon.	Lat.	Lon.
2	43.3553	−5.9228	43.355504	−5.922317	43.3552794	−5.92279815
4	43.35559	−5.92282	43.355448	−5.92277	43.3555372	−5.92282117
6	43.35551	−5.92259	43.355452	−5.92317	43.3554538	−5.92256154
7	43.35548	−5.92285	43.355305	−5.922379	43.3554595	−5.9229905
8	43.35537	−5.92261	43.355607	−5.922746	43.3553385	−5.92265708
10	43.3554	−5.92272	43.355309	−5.922968	43.355372	−5.92270413
11	43.35558	−5.92266	43.355275	−5.922713	43.3556622	−5.92258025
12	43.35551	−5.92277	43.355601	−5.922512	43.3554924	−5.92274692
18	43.35544	−5.92253	43.355435	−5.922579	43.3553941	−5.92255854
19	43.3554	−5.92284	43.355575	−5.922977	43.3553482	−5.92288892



(**a**) (**b**) (**c**)

Figure 16. (**a**) Centroids calculated using the clustering-based method. Each color corresponds to a cluster per player with its centroid marked by the red x. (**b**) Centroids calculated using the centroid method. Each point corresponds with a player (**c**) Centroids calculated using the proposed method. Each point corresponds with a player. All figures refer to matchday 12.

Table 6. Mean of the deviation of the distances calculated for each player using different methods during the first eight matchdays.

Player Number in the Scheme	Player ID Number	Distance Calculated Using the Clustering-Based Method	Distance Calculated Using the Centroid-Based Method	Distance Calculated Using the Proposed Method
2	2	24.38	17.41	19.37
3	4	15.24	20.68	12.63
4	19	8.49	18.50	13.93
5	7	11.54	18.50	8.27
6	10	11.97	12.41	11.75
7	8	13.93	8.71	15.24
8	12	13.71	10.01	10.67
9	18	1.31	14.37	6.97
10	6	4.79	5.01	6.09
11	11	10.45	7.18	6.53
	SUM	115.80	132.78	111.45
	MEAN	11.58	13.28	11.14
	SD	6.24	5.43	4.31

As can be seen in Figure 16c, the positions calculated for the players are closer to reality (the 1-4-2-3-1 tactical scheme showed at Figure 12) than those in Figure 16a,b. Indeed, certain anomalies can be observed in the positioning of some players (two players too close together), which allows undesired behavior to be improved in future situations. As shown in Table 6, although the sum of the differences in the distances of all players with respect to their theoretical position is quite close between the clustering-based method (115.80 m) and the proposed method (111.45 m), the standard deviation is lower by almost 2 m (6.24 m versus 4.31 m), improving the results obtained.

On the other hand, the overview shown by the comparison between the centroid-based method and the proposed method (e.g., the player with ID = 8) allows us to objectively detect individual anomalies in tactical performance. In this case, the distance obtained with the proposed method (15.24 m) is considerably worse than the result of the centroid-based method (8.71 m), which indicates that this player tends to deviate from her zone. This is due to the bias produced by considering only the densest zones when calculating the new centroid.

7. Conclusions and Future Work

As we have seen, there are different methods and algorithms to try to infer the individual roles of each player from the spatio-temporal variables. In this work, a dataset has been elaborated for different categories of women's soccer, from U-12 to the professional category. Some methods used in men's teams have been applied to women's soccer, on the generated dataset, using different techniques: clustering-based methods and methods based on the calculation of centroids. In the first case, the method with which the best results were obtained was the one based on K-means, although the results were not entirely satisfactory. Some authors have refined these methods, showing that better results could be achieved [11]. In the case of centroid-based methods, it has been shown that they prove to be a reliable tool in the determination of the individual role. Finally, a new method based on image analysis using computer vision algorithms has been suggested. Starting from the heat maps generated from the dataset and subjecting them to a previous refinement step, through the detection of hot spots in the image, it is possible to determine the individual role of each player. As shown in Section 5.3, it is possible to apply computer vision algorithms to images to obtain the individual position of the player throughout a match and, therefore, to obtain the tactical scheme of the team, which is the objective of this work. As demonstrated in Section 6, the results improve for those obtained by applying other algorithms. Likewise, the improvements made to the algorithms based on geographical coordinates also improve their efficiency. Since the algorithm only must extract the n hottest areas of the image, it is possible to simplify the input dataset, obtaining an equivalent dataset up to 80% smaller without affecting the final result. For computational purposes, this translates into a performance improvement of close to 90%. On the other hand, the proposed method improves by almost 2 m the calculation of the typical position for each player. This makes it possible to determine more accurately their location on the field and, therefore, also makes it easier to determine the tactical scheme of the team. Also, the fact that we are able to compare the results obtained with the different methods helps us to detect anomalies in both individual and collective behavior, as each method provides different views.

It remains to extend these results and verify their effectiveness on more complete datasets. It could also be useful to test other suggested methods and algorithms or refine some of them to obtain better results.

From the spatio-temporal variables, it would also be possible to predict the time trajectories of each player for better decision making by applying Machine Learning techniques. This would allow an analysis of both the team itself and the opponent. Also, by combining these variables with the collection of critical events that occur during matches, applying artificial intelligence techniques, it would be possible to determine the weaknesses and strengths of the team. Likewise, from the dataset prepared, numerical metrics could be extracted (distances between players of the same line, distances between lines, etc.), which,

related to the critical events, could show situations to be corrected in the team's collective performance. This would provide valuable information to the coaching staff, allowing them to work on these weaknesses in order to eliminate them. This would lead to better results in competition.

Author Contributions: Conceptualization, L.Á.O.; methodology, L.Á.O.; software, L.Á.O.; validation, L.Á.O., D.M. and R.G.; formal analysis, L.Á.O.; investigation, L.Á.O.; resources, L.Á.O.; data curation, L.Á.O.; writing—original draft preparation and editing, L.Á.O.; review D.M. and R.G.; supervision, D.M. and R.G. All authors have read and agreed to the published version of the manuscript.

Funding: This study received no external funding.

Data Availability Statement: Data supporting reported results can be found via this link: https://zenodo.org/records/10913119, accessed on 8 May 2024 [28].

Acknowledgments: The authors would like to thank Kevin Morán Santamarta (Football Physical trainer) and the Gijón women's soccer club, without whose collaboration this work would not have been possible.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ali, A. Measuring soccer skill performance: A review. *Scand. J. Med. Sci. Sports* **2011**, *21*, 170–183. [CrossRef] [PubMed]
2. Koltai, M.; Wallner, D.; Gusztafik, Á.; Sáfár, Z.; Dancs, H.; Simi, H.; Hagenauer, M.; Buchgraber, A.M. Measuring of sport specific skills of football players. *J. Hum. Sport Exerc.* **2022**, *11*, S218–S227. [CrossRef]
3. Ramos, V.H.D.; Román, M.R.; Triguero, D.M.; Godoy, S.J.I.; Lopez, P.S. Relación de la carga de entrenamiento con las emociones y el rendimiento en baloncesto formativo (Relation of training load with emotions and performance in formative basketball). *Retos* **2020**, *40*, 164–173. [CrossRef]
4. Vargas, J.M.G.; Pérez, J.M.G. Análisis descriptivo de variables de rendimiento físico en un equipo de fútbol de primera división chilena femenina (Descriptive analysis of physical performance variables in a Chilean women's first division football team). *Retos* **2023**, *48*, 657–666. [CrossRef]
5. Rico-González, M.; Gómez-Carmona, C.D.; Rojas-Valverde, D.; Los Arcos, A.; Pino-Ortega, J. Electronic Performance & Tracking Systems (EPTS): Practical Applications in Team Sports. In *Libro de Resúmenes del I Congreso Internacional de Iniciación a la Investigación en Ciencias de la Actividad Física y el Deporte*; COLEF Región de Murcia: Murcia, Spain, 2019.
6. Oliva Lozano, J.; Rago, V. *El Sport Scientist y la Monitorización de la Carga con EPTS en Deportes de Equipo*; Editorial Universidad de Almería: Almería, Spain, 2020; ISBN 978-84-1351-041-5.
7. Oliva-Lozano, J.M.; Muyor, J.M. Understanding the FIFA quality performance reports for electronic performance and tracking systems: From science to practice. *Sci. Med. Footb.* **2021**, *6*, 398–403. [CrossRef]
8. Memmert, D.; Lemmink, K.A.P.M.; Sampaio, J. Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Med.* **2016**, *47*, 1–10. [CrossRef]
9. FIFA. Informe Comparativo de Fútbol Femenino. Versión Español. 2019. Available online: https://digitalhub.fifa.com/m/31350ff23e84e0fa/original/Informe-de-evaluacion-comparativa-de-la-FIFA-futbol-femenino.pdf (accessed on 7 May 2024).
10. Bialkowski, A.; Lucey, P.; Carr, P.; Yue, Y.; Sridharan, S.; Matthews, I. Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data. In Proceedings of the 2014 IEEE International Conference on Data Mining, Shenzhen, China, 14–17 December 2014.
11. Behravan, I.; Zahiri, S.H.; Razavi, S.M.; Trasarti, R. Finding Roles of Players in Football Using Automatic Particle Swarm Optimization-Clustering Algorithm. *Big Data* **2019**, *7*, 35–56. [CrossRef]
12. Clemente, F.M.; Couceiro, M.S.; Martins, F.M.L.; Mendes, R.S.; Figueiredo, A.J. Intelligent systems for analyzing soccer games: The weighted centroid. *Ing. Investig.* **2014**, *34*, 70–75. [CrossRef]
13. Folgado, H.; Lemmink, K.A.P.M.; Frencken, W.; Sampaio, J. Length, width and centroid distance as measures of teams tactical performance in youth football. *Eur. J. Sport Sci.* **2012**, *14* (Suppl. 1), S487–S492. [CrossRef] [PubMed]
14. Castellano, J.; Figueira, B.; Coutinho, D. Identifying the effects from the quality of opposition in a football team positioning strategy. *Int. J. Perform. Anal. Sport.* **2013**, *13*, 822–832. [CrossRef]
15. Sampaio, J.; Maçãs, V. Measuring tactical behaviour in football. *Int. J. Sports Med.* **2012**, *33*, 395–401. [CrossRef] [PubMed]
16. Fonseca, S.; Milho, J.; Travassos, B.; Araújo, D. Spatial dynamics of team sports exposed by Voronoi diagrams. *Hum. Mov. Sci.* **2012**, *31*, 1652–1659. [CrossRef] [PubMed]
17. Gudmundsson, J.; Wolle, T. Towards Automated Football Analysis: Algorithms and Data Structures. In Proceedings of the 10th Australasian Conference on Mathematics and Computers in Sport, Darwin, Australia, 5–7 July 2010.

18. Wei, X.; Sha, L.; Lucey, P.; Morgan, S.; Sridharan, S. Large-Scale Analysis of Formations in Soccer. In Proceedings of the 2013 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2013, Hobart, Australia, 26–28 November 2013; pp. 1–8. [CrossRef]
19. Linke, D.; Lames, M. Validation of electronic performance and tracking systems EPTS under field conditions. *PLoS ONE* **2018**, *13*, e0199519. [CrossRef]
20. Rodríguez, L.A.O.; Fernández, R.G. Low cost EPTS (Electronic Performance & Tracking System) development using IoT devices. In Proceedings of the 2021 16th Iberian Conference on Information Systems and Technologies (CISTI), Chaves, Portugal, 23–26 June 2021; pp. 1–4. [CrossRef]
21. OHCOACH Help. Available online: https://help.ohcoach.com/ (accessed on 7 May 2024).
22. Walter, F.; Lames, M.; McGarry, T. Analysis of sports performance as a dynamic system by means of relative phase. *Int. J. Comput. Sci. Sport* **2007**, *6*, 35–41.
23. Olthof, S.B.; Frencken, W.G.; Lemmink, K.A. The older, the wider: On-field tactical behavior of elite-standard youth soccer players in small-sided games. *Hum. Mov. Sci.* **2015**, *41*, 92–102. [CrossRef] [PubMed]
24. Gonçalves, B.V.; Figueira, B.E.; Maças, V.; Sampaio, J. Effect of player position on movement behaviour, physical and physiological performances during an 11-a-side football game. *J. Sports Sci.* **2014**, *32*, 191–199. [CrossRef] [PubMed]
25. Frencken, W.; Lemmink, K.; Delleman, N.; Visscher, C. Oscillations of centroid position and surface area of soccer teams in small-sided games. *Eur. J. Sport Sci.* **2011**, *11*, 215–223. [CrossRef]
26. Bialkowski, A.; Lucey, P.; Carr, P.; Denman, S.; Matthews, I.; Sridharan, S. Recognising team activities from noisy data. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013; pp. 984–990.
27. Buchheit, M.; Allen, A.; Poon, T.K.; Modonutti, M.; Gregson, W.; Di Salvo, V. Integrating different tracking systems in football: Multiple camera semiautomatic system, local position measurement and GPS technologies. *J. Sports Sci.* **2014**, *32*, 1844–1857. [CrossRef] [PubMed]
28. Rodríguez, L.Á.O. Women's team soccer positioning data, collected during 2023/2024 season. Third category of female soccer, Spain. *Zenodo* **2024**, *3*. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.