Initial Approach on Soccer Match's Scene Classification by Players' Field Spatial Distribution

Abordaje inicial en la clasificación de escenas de partidos de fútbol a partir de la distribución espacial de los jugadores sobre la cancha

Lennon Núñez-Meoño[1], Marco Villalta[2], Francisco Siles-Canales[3]

Núñez-Meoño, L; Villalta, M; Siles-Canales, F. Initial Approach on Soccer Match's Scene Classification by Players' Field Spatial Distribution. *Tecnología en Marcha*. Edición especial 2020. 6th Latin America High Performance Computing Conference (CARLA). Pág 60-65.

https://doi.org/10.18845/tm.v33i5.5077

1 PRIS-Lab: Pattern Recognition and Intelligent Systems Laboratory, Department of Electrical Engineering, School of Engineering and Postgraduate Studies in Electrical Engineering, Universidad de Costa Rica (UCR). CNCA: National Advanced Computing Collaboratory, CeNAT: National Center for High Technology. Costa Rica. E-mail: lennon.nunez@ucr.ac.cr
2 PRIS-Lab: Pattern Recognition and Intelligent Systems Laboratory, Department of Electrical Engineering, School of Engineering and Postgraduate Studies in Electrical Engineering, Universidad de Costa Rica (UCR). E-mail: marco.villalta@ucr.ac.cr
3 PRIS-Lab: Pattern Recognition and Intelligent System Laboratory, Department of Electrical Engineering, School of Engineering, Universidad de Costa Rica, Costa Rica. Email: francisco.siles@ucr.ac.cr.
https://orcid.org/0000-0002-6704-0600

Keywords

Soccer player; object tracking; scene classification; occlusion; stretch index; player dispersion; surface area.

Abstract

Soccer player tracking is moving towards the use of high-cost algorithms; however, these are implemented to the whole video recording of interest with no flexibility on the type of scene being evaluated. The current work proposes to classify the scenes according to the players' distribution over the field, designing a metric for players' Dispersion Level that would allow implementing a more flexible tracking method that selects a light algorithm for well-behaved scenes and a heavy and robust algorithm for bad-behaved scenes. This work pushes towards considering different existent metrics, as Voronoi diagrams for surface area analysis, in future works to come.

Palabras clave

Jugador de fútbol; rastreo de objetos; clasificación de escenas; oclusión; índice de estrechés; dispersión de los jugadores; área superficial.

Resumen

El rastreo de jugadores de fútbol se inclina al uso de algoritmos de alto costo computacional, sin embargo, estos suelen ser implementados de manera pareja al video de interés sin consideración de los tipos de escena evaluada. El trabajo presente propone clasificar las escenas con respecto a la distribución espacial de los jugadores sobre el campo por medio de una métrica del nivel de dispersión de los jugadores, con la cual implementar métodos de rastreo flexibles en cuanto a la selección de algoritmos livianos para escenas de 'buen carácter', y algoritmos pesados y robustos para escenas de 'mal carácter'. En investigaciones a venir, se pretende evaluar distintas métricas, como el uso de diagramas de Voronoi para el análisis del área superficial del espacio entre jugadores.

Introduction and Related Work

In many sport disciplines, video recordings are used to extract statistics on performance, the coach or technical director is in charge of studying that information, or even to extract it himself, in order to design the best strategy and help the athletes to improve; modern techniques include algorithms on object tracking to automate this process. In the case of soccer teams, players tracking provides information of the position of the player at any time during the whole match, with it, it is possible to tell different performance statistics: total time played, distance traveled, fatigue and placement of the player during different key plays [8]. Thus, players' performance and tactical evaluation can be assessed based on information from object tracking, making it a valuable tool for sports analysis.

When implementing multiple object tracking, one of the biggest problems is the occlusion event, especially when using a single point of view, as it only requires two overlapping players for it to happen. Different algorithms have been developed to solve this issue [3,1], commonly with high computational cost, the common approach is to apply these algorithms to the whole video recording, usually distributing the frames into different blocks that are merged to give the final result, there is no decision over which algorithm is best for a specific scene as there is no current information of what's going on until the tracking results are obtained. Many platforms rely on the human interpretation of the different event scenes to classify them by game plays, as automatically semantic interpretation is still in early development [5]. If the scenes were automatically classified before implementing the tracking algorithm, different decisions could be made according to the scene type, as when to use a heavy algorithm to treat occlusion cases or when to use a light algorithm for 'Well Behaved' scenes.

Automatic scene classification started as a method to aid the tracking on sports broadcasting, where there are many changes in perspective (camera changes, zoom-in, and zoom-out) as well as sequence loss (commercials, replays) without prior knowledge of when any of these will occur. The broadcasting methods rely on color-based filtering to recognize the scenes [10,7], where there is a noticeable difference in color properties. In an Ultra High Definition recording of the whole field, all these special cases are non-existent as the view is maintained throughout the match; there is no need to classify scenes into zoom-in or zoom-out, commercials or match events.

As mentioned for the examples above, trends focus on broadcast videos, therefore, the scene classification types are accordingly, separating the events as shots. [9] classifies the shots in classes as transitional effects with different span times, overlays, long-distance and close up shorts. However, the proposed scene classification focuses on acquiring the ability to detect possible occlusion events. On this topic, [7] implements the Occlusion Alarm Probability (OAP) to detect players' proximities. This solution is an approach into detecting occlusion events, but there is not a classification or indexing of the players proximity events.

Currently, scene classification by game plays is performed fully manually or human aided, moreover, in occlusion matters, the scenes are identified as occlusion risky scenes just after the tracking information is obtained, as it is easier to detect when there's a blob merge or split. Therefore, the aim of the current effort is to design a metric for scenes evaluation prior to obtaining the tracking results, with which to classify the game match scenes by players congestion and distribution over the soccer field, classification that will aid decision making on different tracking approaches based on the type of scene in process, specially to decide between a light or a heavy algorithm if there is a high chance of occlusion.

Methodology

The occlusion events occur as the players get closer to each other, therefore, finding a metric that can numerically represent the distribution of the players over the game field can help to implement flexible tracking methods depending on the agglomeration complexity of the scene, or, in other words, the players dispersion level over the field. As the event to measure is agglomeration, i.e. how close the players are from each other, distance player to player is evaluated. Different alternatives are considered, [2,4] mention three main metrics, Stretch Index, Surface Area, and Team Length. The Stretch Index compares the average distance of a team's player and geometrical center, being a measure of how compact or stretched is the team. Taking $P_{axis.n}(k)$ as the n'th player's position on $axis = X, Y$ on the frame k, $GC_{axis}(k)$ the position of the geometrical center of the team on frame k, and N the total number of team players on the frame k, the Stretch Index (SI) is calculated per equation 1.

$$SI = \frac{\sum_{n}^{N} \sqrt{\left(P_{X,n}(k) - GC_X(k)\right)^2 + \left(P_{Y,n}(k) - GC_Y(k)\right)^2}}{N} \quad \text{Equation 1}$$

The equation for *SI* is, therefore, the average of magnitudes of the vectors *cluster center to player* for each player of the current team. This metric is designed taking into account that the players are separated into two teams, we intend to see all the players as a whole single group, as occlusion may occur with players from opposing teams, thus, calculating a distribution of all the individuals over the field is required.

A useful metric for measuring dispersion is the standard deviation, its value can help to assess how close the players are in comparison to a common reference, an approach similar to SI with the difference that it considers the average of the measurements as the reference for dispersion. Given the frame independent reference *R* with position (X, Y), the vector player-to-reference $PR_n(k) = P_n(k) - R$, with $P_n(k)$ the player's position from the origin of the image's frame, and the mean of the vector's magnitudes μ, the standard deviation σ of the frame k for a single reference is calculated by the equation 2.

$$\sigma = \sqrt{\frac{1}{N}\sum_{n}^{N}(PR_n(k) \vee -\mu)^2}$$
Equation 2

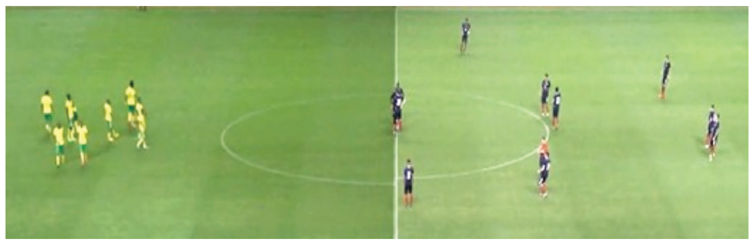
Considering that SI is the average of magnitudes with a specific reference (*GC(k)*), the standard deviation can be described by *SI(R)* with *R* as the desired reference, taking the equation 2 and replacing μ with *SI(R)*. We take into account four different references (each corner of the game field) and calculate the standard deviation for each, the result is four standard deviations.

Experiments and Preliminary Results

One soccer match was picked from the PRIS-Lab's video database, selected by the quality of the panoramic take (middle line alignment, view perspective), from which the scenes with a more noticeable change on dispersion are manually picked for the experiment. The four edges of the game area are used as references. Making the supposition that players tend to be somehow dispersed over the field, a trend is calculated to define the ranges of a Dispersion Level (DL), this to classify the scenes in three classes that would tell if the frames have a high probability to present occlusion: High, Medium and Low Level of dispersion, or, Well, Medium Well and Bad-behaved scenes, respectively. It is expected that for the frames where the players are visually close to each other, the standard deviation is lower than in those frames where the players are spaced or spread along the field. The above assumption can be observed in figure 1, which describes three main scenarios: the match starts with a big players agglomeration (low standard deviation, bad-behaved scene), the players start to disperse but all move following the ball (standard deviation increases, medium well-behaved scene), and the players space along the field (standard deviation reaches the highest values, well-behaved scene). The standard deviation for the first 41 seconds (in frames) of the match using the four corners of the field as references is shown in figure 2. The first 300 frames of video (lowest standard deviation) correspond to Bad-Behaved scene (figure 1.c), the middle of the match (frames about 300 to 900) are Medium Well Behaved scene (figure 1.b), and last frames are Well Behaved scene (Figure 1.a) when ball gets out of play and players disperse back to the center.

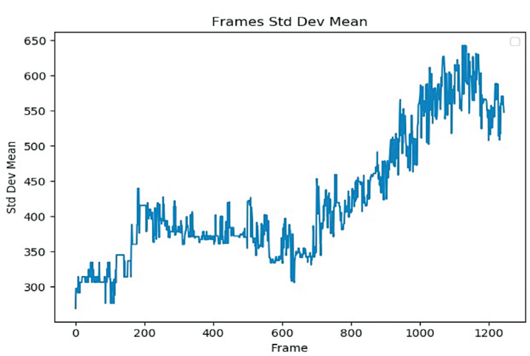
(a) Well Behaved scene

(b) Medium Well Behaved Scene

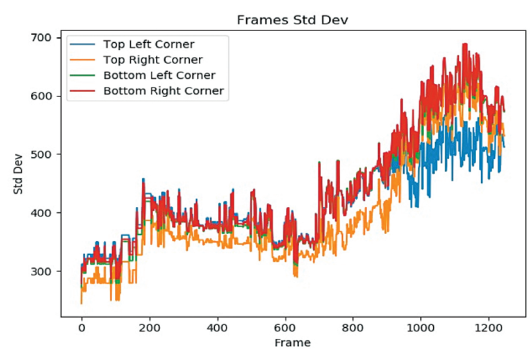


(c) Bad-Behaved scene

Figure 1. (a) Players are spread along the field. (b) Players are spread along the field but closer to each other. (c) Players are agglomerated in different clusters.



(a) Mean of standard deviations



(b) Standard deviation for each reference

Figure 2. Standard Deviation of all the players on a soccer match using the four corners of the soccer field as reference

Conclusions and future work

The presented metric is a simple approach independent from the object tracking results, as it requires only the blob's 2D positions, regardless of which player is which or to which team it corresponds. However, more experiments are required to provide a quantitative result on its capacity to properly distinguish the required scenes, including longer video clips and more study cases. Also, as the tracking information is not used, spurious blobs will affect the calculations, a situation that would also affect even having the tracking information as a no-error fully automatic tracking platform doesn't exist yet. This metric cannot tell if there is or not occlusion on the image, it gives a hint on which frames have a high probability of presenting tracking issues (mostly occlusion) due to a low players' Dispersion Level; even in a Well Behaved scene, there could be a two players occlusion event. Thus, the need for testing different metrics to obtain a more robust method to detect these special cases. One consideration is to use surface area metrics based on Voronoi diagrams, with which to evaluate the size of the area spaces between players inside the soccer field, the smaller the area in between, the more agglomerated the players are.

References

[1] Amditis, A., Thomaidis, G., Maroudis, P., Lytrivis, P., Karaseitanidis, G. (2012). Multiple Hypothesis Tracking Implementation, chap. 10. Laser Scanner Technology, IntechOpen.

[2] Clemente, M., Sequeiros, J., Correira, A., Sylva, F., Laurenço, F. (2012). Computational Metrics for Soccer Analysis: Connecting the dots, chap. 5. SpringerBriefs in Applied Sciences and Technologies, Springer.

[3] Dearden, A., Demiris, Y., Grau, O. (2006). Tracking football player movement form a single moving camera using particle filters. In: CVMP-2006. pp. 29–37.

[4] Folgado, H., Lemmink, K., Frencken, W., Sampaio, J. (2014). Length, width and centroid distance as measures of teams' tactical performance in youth football. European Journal of Sport Science (14), pp. 487–492.

[5] Kapela, R., McGuinness, K., O'Connor, Noel E. (2017). Real-time field sports scene classification using colour and frequency space decompositions. Journal of Real-Time Image Processing 13(4), 725–737.

[6] Mei, T., Ma, Y.-F., Zhou, H.-Q., Ma, W,-Y., Zhang, H.-J. (2005). Sports video mining with mosaic. In: 2005 International Conference on Multimedia Modelling. pp. 107–114.

[7] Ok, H.-W., Seo, Y., Hong, K.-S. (2002). Multiple soccer players tracking by condensation with occlusion alarm probability. Tech. rep., IIP Lab., Pohang University of Science and Technology (POSTECH), Republic of Korea.

[8] Radakovic, R. Dopsaj, M. Vulovic, R. (2015). The reliability of motion analysis of elite soccer players during match measured by the tracking motion software system. In: 2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE).

[9] Siles, F. (2014). Automated Semantic Annotation of Football Games from TV Broadcast. Ph.D. thesis, Technische Universitad Munchen.

[10] Zhong, D., Chang, S. (2004). Real-time field sports scene classification using color and frequency space decompositions. Journal of Visual Communication and Image Representation 15(3), pp. 330–347.