Head, Shoulders, Hip and Ball... Hip and Ball! Using Pose Data to Leverage Football Player Orientation

Adrià Arbués-Sangüesa, Gloria Haro, Coloma Ballester, Adrián Martín

Universitat Pompeu Fabra | adria.arbues@upf.edu

Abstract

Orientation has proven to be a key skill of football players in order to succeed in a broad spectrum of plays, such as a receiving or giving a pass because of an appropriate field of view, defending two players at a time or finding open-spaces due to a fast reaction. However, body orientation is a yet-little-explored area in sports analytics. By seeking the 2D orientation of the field projection of the normal vector placed in the center of the upper-torso of players, this research presents a novel technique to extract orientation automatically from video recordings by merging pose and contextual information. On the one hand, OpenPose is used in combination with a super-resolution network to extract the coordinates of body parts of every single player; by mapping pose parts (shoulders and hips) in a 2D field, a first orientation estimation is obtained; on the other hand, contextual information quantifies the orientation of each player with respect both to the ball and to the field position. Results have been validated with players-held EPTS devices, obtaining a 92.4% accuracy on left-right orientation directions and an absolute median error of less than 35 degrees/player. A handful of practical applications is provided, from improving and refining pass probability and pitch-control models, to producing a fine-grained evaluation of player's on-ball and off-ball direction and turning speed.

1 Introduction

The recent rise of sports analytics has provided a new set of metrics and statistics that can serve coaches to evaluate both player's and team performance. From spatio-temporal models that estimate the probability of possession success in football [4], to the forecasting of future movement locations in basketball [8] or advanced statistics used in curling [6], tracking data has provided a rich source of information for exploring complex spatio-temporal dynamics in team sports. Despite their importance, these location data are clearly insufficient to determine if a player is in condition of properly acting during the play, which will be influenced by contextual information and the player's own pose and orientation. Complex metrics could be built from player orientation, which is a critical piece of information that remains unquantified. The main goal of this research is to estimate the body orientation of football players from video data, with potential generalization to other sports.

Obtaining a time-based set of orientations would not only imply an improvement of football analytics itself, but also improving and refining pass probability and pitch-control models, as well as producing a fine-grained evaluation of player's on-ball and off-ball direction and turning speed. Moreover, its integration with video allows this model to be used as a coaching tool to assess players' orientation under different situations.

In this paper, player orientation is gathered by combining three different types of estimations: pose, field and ball orientation. The model's outcome has been validated with a ground-truth orientation dataset, gathered with EPTS devices in football games of F.C. Barcelona B team, obtaining promising results. Different tests show that: (a) obtained orientation faces the proper side of the court (right/left) in 92.4% of the cases, and (b) the median absolute error is around 34 degrees/player.

2 Methods

In this work, the orientation of a player's body is defined as the 2D rotation of the player's upper-torso around the vertical axis, which is assumed to coincide with the field projection of a normal vector placed in the center of their upper-torso, involving both shoulders and hip parts. The overall pipeline of the presented method is displayed in Figure 1. This Section provides a detailed explanation of 3 different kinds of orientation estimation from which the method benefits: (1) pose data, (2) field position and (3) ball position. The output of all these individual estimations produces both a numerical orientation result and a confidence value. Orientation is measured in degrees and discretized into 24 probability bins using the reference system displayed in Figure 3(a). For all cases, the orientation value indicates the bin with higher probability, while the confidence value is used to compute how many other neighboring bins have non-zero probability. This paper proposes an algorithm that outputs a probability density function (pdf) of the estimated orientation; thus containing both an estimation (maximum of pdf) and its confidence (inverse of the pdf support). *A posteriori*, a contextual weighting is performed.

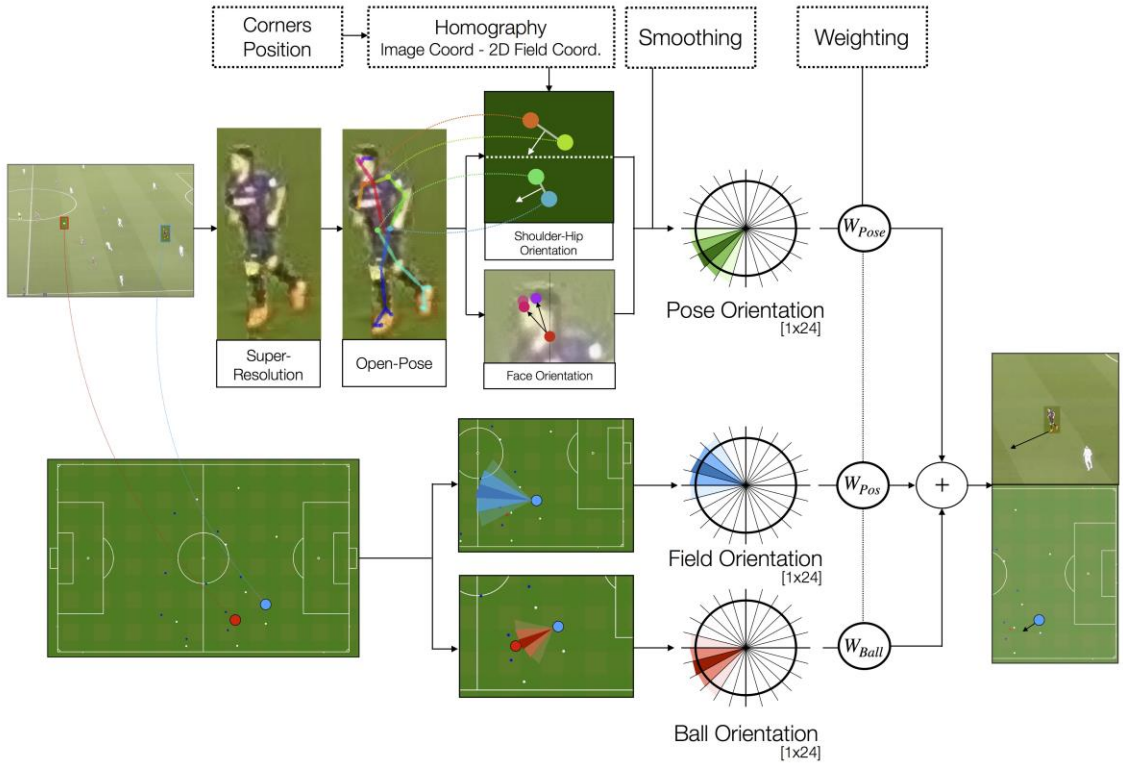

Figure 1. Proposed pipeline. On the one hand, pose orientation is found by combining a super-resolution network, OpenPose and 3D vision techniques; on the other hand, field orientation and ball orientation are also computed. A combination is later performed.

2.1 Pose Orientation

Orientation from pose is the main contribution of this article, and uses pretrained models and 3D vision techniques in order to obtain a first orientation estimation of each player. Given temporally-smoothed bounding boxes of players, a combination of super-resolution and pose detection techniques is applied to find the pose of every player. Both the left-right shoulders and the left-right parts of the hip will be considered as the main upper-torso parts. By projecting

these parts in a 2D space independently, the normal vector between these points can be extracted.

2.1.1 Pose Detection

Having the bounding boxes for all visible players in each frame, the OpenPose library [1] is used to extract the pose of every single individual (see [7,9,2] for technical details of pose models). Given a football frame, the output of the pose estimator is a 25 x 3 vector for each player, with the position (in image coordinates) of 25 keypoints, which belong to the main biometric human-body parts, together with a confidence score.

However, detecting the pose of players in sports scenarios is always challenging given the frequent occlusions and fast movements that lead to motion blur. Moreover, football games are played in fields of approximately 7,700 squared meters, so depending on the video footage, players might result in tiny portions of the image; for instance, given a set of 3,047,270 bounding boxes in Full-HD images, the mean size of a single bounding box has a resolution of only 14x49 pixels. Hence, small image crops are not always properly processed, resulting in a null set of anatomical landmarks. For this reason, a super-resolution network is used to preprocess bounding boxes and enhance the image quality instead of a simpler interpolation technique. More concretely, the applied model is a Residual Dense Network (RDN)[3,10]. The resulting processed image is 4 times bigger than the original in width and height. In 2, an example of (a) the original low resolution image can be seen together with upscaled images obtained using (b) bicubic interpolation and (c) the implemented RDN. The image processed by the RDN is less blurry and allows OpenPose to better detect the body parts of the players. However, two other challenges remain unsolved:(1) double detections, solved by comparing histograms based on given templates, and (2) non-existent detections, which might happen either by an inaccurate smoothing when tracking players due to sudden changes or a bad performance of OpenPose in the given crop.

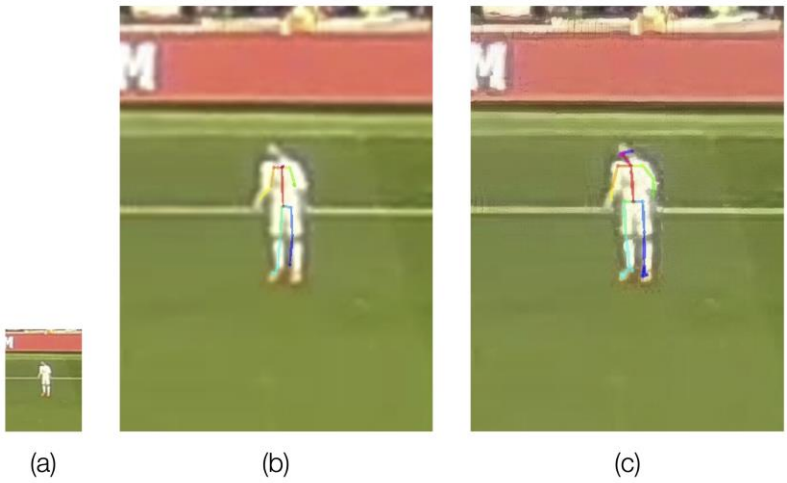

Figure 2. Super-resolution performance: (a) original cropped image according to bounding box coordinates (plus a reasonable margin). OpenPose result after upscaling with a: (b) bicubic interpolation, and (c) a Dense Residual Network [10].

2.1.2 Angle Estimation

Once the pose is extracted for each player, the coordinates (and confidence) associated to the upper-torso parts are stored to estimate the pose orientation. Given the 4 field corners' coordinates in the image (or its projection out of the image in the non-visible cases), a homography is computed with the DLT algorithm [5] after establishing the four field-corners correspondences between the frame and a 2D field given by a template image of it. This homography is a non-singular 3x3 matrix that maps field-pixels seen in the frame domain into 2D field coordinates. From the output of OpenPose, the coordinates of the main upper-torso parts are found in the image domain; by mapping the left-right pair (either shoulders or hips) in the 2D field, a first insight of the player orientation is obtained, as seen in Figure 3(b). Basically, the player can be inclined towards the right (0-90º, 270-360º, bins 0-11) or the left (90-270º, bins 12-23) side of the field. Extreme cases, such as top (90º) or bottom (270º), are also considered. From now on, this first binary estimation, which indicates if the orientation belongs to the first or second half of the orientation histogram, will be called *LR-side* parameter.

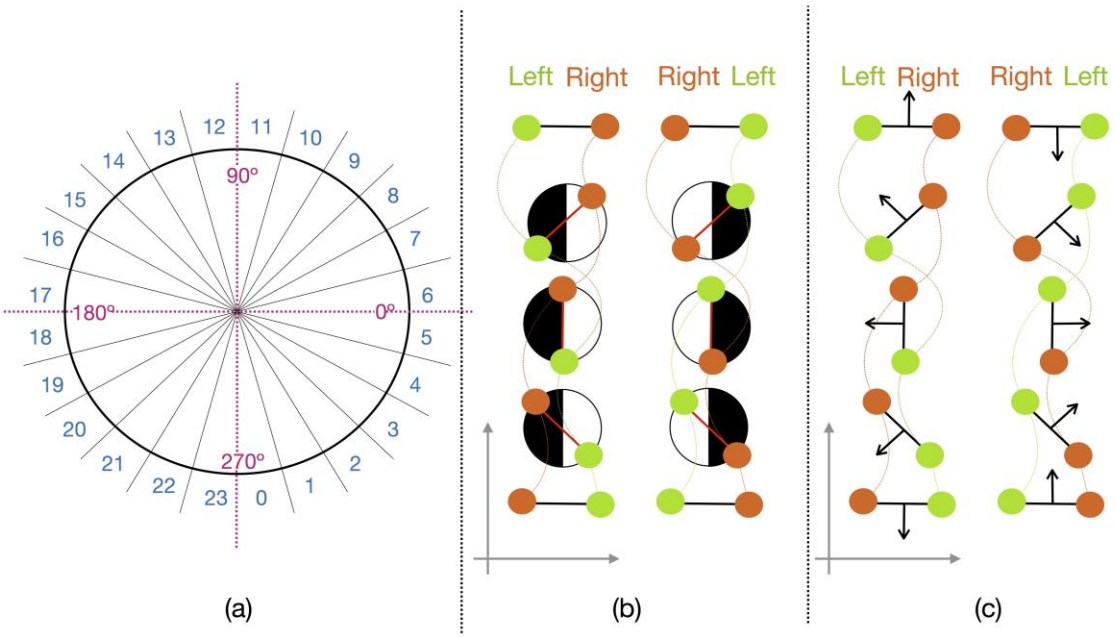

Figure 3. (a) Orientation discretization in 24 bins; blue numbers stand for the bin number and purple angles stand for the related orientation. (b) Different combinations of left-right mapped parts in a 2D space, showing if the player's orientation belongs to the first or second half of the histogram (black semicircle, LR-side parameter); (c) same combinations with the corresponding normal vector.

Figure 4 shows how pose orientation is estimated: first, left-right shoulders and hips are mapped via the estimated homography into the 2D space; then, *LR-side* booleans (LR_{Sh}, LR_{Hi}), angles (α_{Sh}, α_{Hi}) and confidences (C_{Sh},C_{Hi}) are obtained. The associated confidences are the product of OpenPose's individual shoulder and hips confidences respectively. However, OpenPose might fail detecting either the left or right hip parts; in these cases, the middle hip position is used as a substitute for the missing part. Then:

1. If LR_{Sh} and LR_{Hi} agree:
    a. If $C_{Sh} > C_{Hi}$, α_{Sh} is considered as the pose orientation estimation and C_{Sh} its confidence. If not, α_{Hi} and C_{Hi} are selected.
2. Otherwise, there are two possible scenarios: a low confidence detection of some individual parts, or a player placed completely parallel to sidelines with both shoulders and hip parts located in very close pixels. Therefore, confidence values and a computation of the face direction are used:
    a. If $|C_{Sh} - C_{Hi}|$ is bigger than a threshold (*i.e,* 0.4 for both $C_{Sh}, C_{Hi} \in [0,1]$), it means that OpenPose's result is uncertain in one of these parts so only the other one is used for the estimation.
    b. If $|C_{Sh} - C_{Hi}|$ is smaller than this threshold, the player's face direction is checked. In the image domain, the difference among the X positions of all face parts and the player's neck is computed. If most of the parts move towards the origin of the X axis (Figure 4(c)), the player's *LR-side* will be left, and the angle 180; otherwise, the player's *LR-side* will be right and the angle 0.

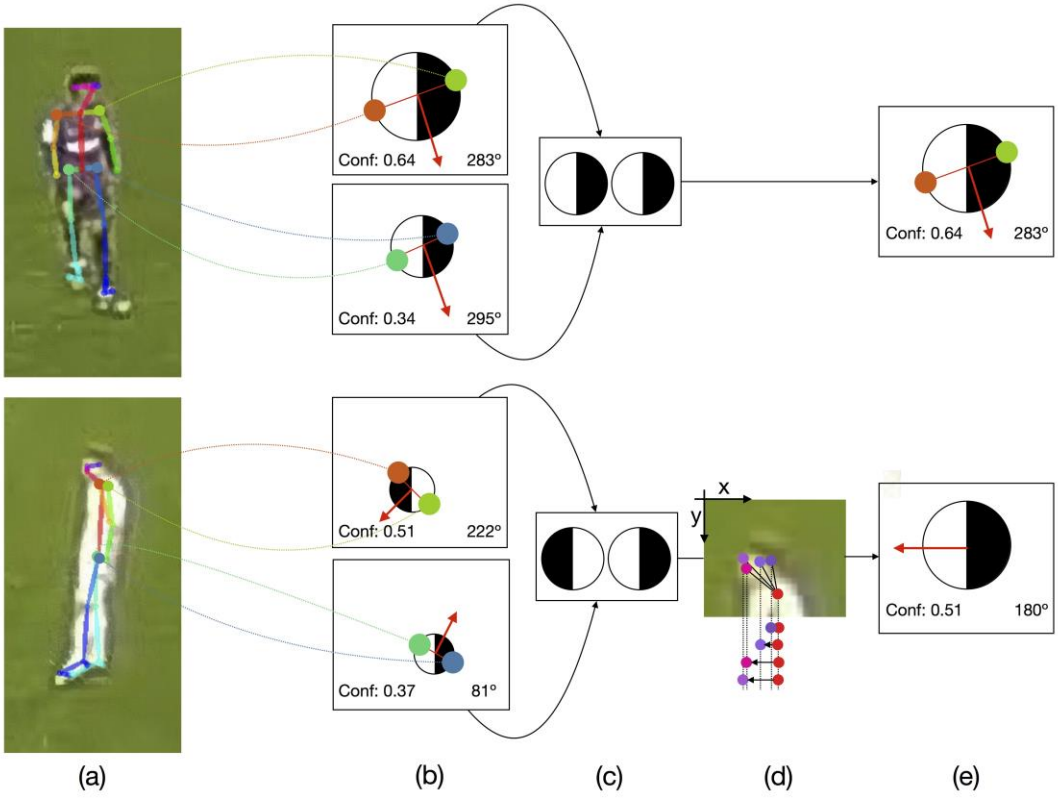

Figure 4. Steps to estimate the pose orientation of a given player and its LR-side parameter: (a) OpenPose output, (b) coordinates belonging to shoulders and hip are mapped in a 2D field, individual LR-side and orientation values are obtained. (c) A simple check is performed to see if both LR-parameters coincide; if they do not and their confidence difference is below a certain threshold, (d) the face direction is checked by comparing the X coordinate of all face parts with the neck; (e) a final estimation is obtained according to: (top) higher confidence value, or (bottom) face direction.

Then, given the final pose orientation estimation α_P and its related confidence C_P, a Gaussian probability distribution is located around it, with effective support

$$N_P = \max\left(\left\lfloor N_{bins}\left(\frac{1-C_P}{2}\right)\right\rfloor, 1\right),$$

centered at

$$\mathrm{or}_P = \begin{cases} \left\lfloor \frac{\alpha_P}{360/N_{bins}} + \frac{N_{bins}}{4} \right\rfloor & \text{if } \frac{\alpha_P}{360/N_{bins}} < 18 \\ \left\lfloor \frac{\alpha_P}{360/N_{bins}} + \frac{N_{bins}}{4} \right\rfloor - N_{bins} & \text{if } \frac{\alpha_P}{360/N_{bins}} \geqslant 18 \end{cases}$$

where the second element of the sum is an offset that compensates the bin order (Figure 3(a)). The output vector of this orientation estimation will be denoted from now on H_P. Finally, a median filter in the temporal domain is applied to H_P to correct from inconsistencies in the framewise OpenPose detection.

2.2 Field Orientation

The next estimation to be computed is field orientation, which aims to quantify the field of view of each player at a time. For instance, a player running close to the sideline will rarely be looking outside the field, thus the field of view is reduced, *a priori*, to 90ª according to the associated *LR-side* parameter; likewise, a player right in the center of the field has a broad field of view of 180º, as seen in Figure 5. In middle cases, the field of view changes depending on the distance of the player with respect to the center of the field; the closer the player is to the middle, the wider the field of view. In real cases, players move freely, and their field of view is not restricted, so this estimation, rather than providing an accurate result, might help unbiasing other misleading orientations.

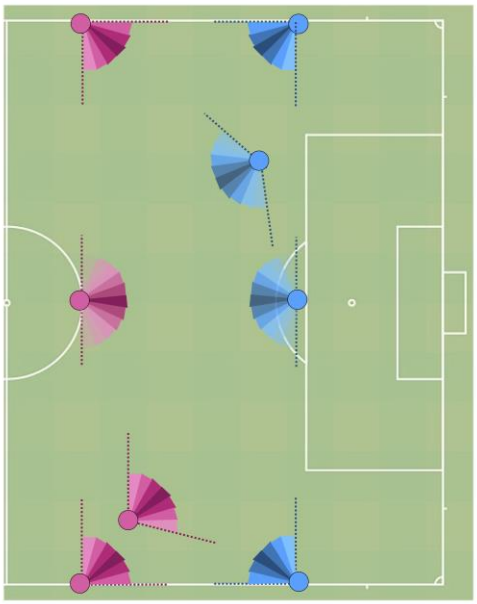

Figure 5. Different types of field of view depending on the individual LR-side; this parameter indicates that 4 players are facing the right side of the field (maroon circles), while the other 4 are facing the left side (blue ones)

Given a player 2D coordinates Pl_X, PL_Y and a 2D field with size (w, h), the effective support of the Gaussian is

$$N_F = \left\lfloor \frac{N_{bins}}{2} - \left| \frac{Pl_y - h/2}{h/2} \right| \cdot \frac{N_{bins}}{4} \right\rfloor$$

and the center depends on the *LR-side* boolean:

$$\text{or}_F = \begin{cases} \frac{Pl_y}{h/6} + \frac{N_{bins}}{6} & \text{if LR-side} = \text{R} \\ \frac{N_{bins}}{2} + \frac{Pl_y}{h/6} + \frac{N_{bins}}{6} & \text{if LR-side} = \text{L} \end{cases}$$

The outcome of this estimation is another discretized probability vector, called from now on H_F.

2.3 Ball Orientation

The last estimation is related to the position of the ball. Logically, players close to the ball tend to be strongly oriented towards it, while players placed far away may not have to be duly oriented accordingly. Hence, having all pairwise distances and the corresponding angles, the orientation of players with respect to the ball can be estimated, as shown in Figure 6.

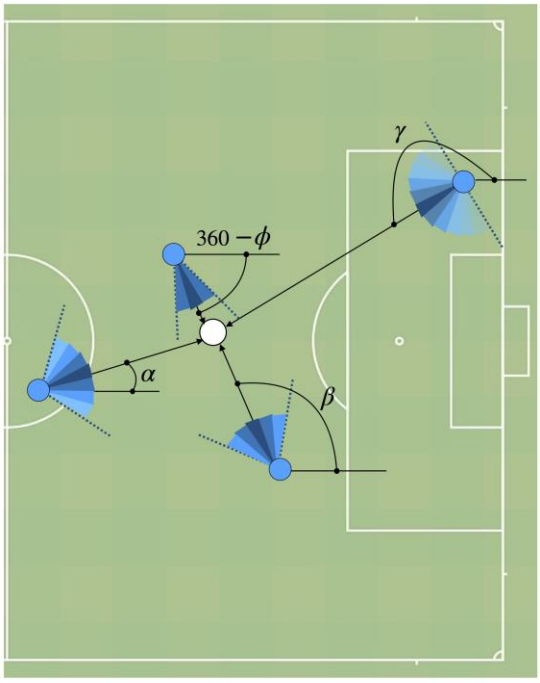

Figure 6. Orientation computation with respect to the ball of 4 different players, considering both the angle (direction) and the distance (magnitude). The same reference system shown in Figure 3(a) is used.

Then, for a given player at (P_X, P_Y), in a moment where the ball is at (B_X, B_Y)generating an angle of βdegrees, the effective support is:

$$N_F = \frac{N_{bins}}{4} \left\lfloor 1 - \frac{MD - \sqrt{(P_x^2 - B_x^2) + (P_y^2 - B_y^2)}}{MD} \right\rfloor + \frac{N_{bins}}{8}$$

where MD is a maximum distance value that regularizes how far a player can be from the ball without being influenced by it (*i.e.* 1/6 of the field's area). Then, the central bin to be filled with the highest weight is:

$$\mathrm{or}_B = \begin{cases} \left\lfloor \frac{\beta}{360/N_{bins}} + \frac{N_{bins}}{4} \right\rfloor & \text{if } \left\lfloor \frac{\beta}{360/N_{bins}} \right\rfloor < 18 \\ \left\lfloor \frac{\beta}{360/N_{bins}} + \frac{N_{bins}}{4} \right\rfloor - N_{bins} & \text{if } \left\lfloor \frac{\beta}{360/N_{bins}} \right\rfloor \geqslant 18 \end{cases}$$

Once again, the outcome of this estimation is a discretized probability vector, called from now on H_B.

2.4 Contextual Merging

Once all three histograms are obtained, a simple weighting is performed among them, thus merging pose, field and ball orientations. In particular, three different weights are used:

$$H_{\mathrm{TOT}} = W_P \cdot H_P + W_F \cdot H_F + W_B \cdot H_B,$$
$$\text{s.t. } W_P, W_F, W_B \in [0,1] \text{ and } W_P + W_F + W_B = 1$$

Nevertheless, the field orientation accuracy depends directly on the *L-R side* parameter, extracted when computing pose orientation. For this reason, both H_P and H_F benefit from pose orientation, hence the relevance of pose information while estimating player's orientation is not only measured with W_P, but also with W_F.

The orientation θ of each player is the central value of the bin H_{TOT} with higher weight:

$$\theta = \mathrm{argmax}\left(H_{\mathrm{TOT}}\right) \cdot \frac{360}{N_{bins}} + \frac{360/N_{bins}}{2},$$

In terms of visualization, orientations can be displayed in the 2D field; starting from a 2D point P_b, which indicates the position of a given player, another point P_e can be projected at a given distance T_θ having an orientation of θ degrees; as a result, the vector joining P_b and P_e will have the estimated orientation and a length proportional to the estimated confidence T_θ. Moreover, having both points, the same coordinates can be mapped back into the original frames by multiplying P_b and P_e by the inverse homography (Subsection 2.1.2), thus showing the orientation vector in the video frame.

3. Results

3.1 Dataset

The dataset provided by F.C. Barcelona included video footage (25 fps and Full-HD) of several games from La Liga, smoothed tracking data in both frame and field domains, corners positions and contextual information. Moreover, data from youth games was gathered, with video footage and raw data extracted from EPTS devices, containing 100 samples/second about player ground-truth XYZ orientation.

Nevertheless, it is worth mentioning that the video quality of third division games is not as neat as desired; plus, given the low height of the stadium, the type of camera shot is prone to cause more occlusions than usual, as seen in Figure 7. Videos have been subsampled to one third of the original frame rate, and a total of 1000 frames and 10395 player instances have been used to validate results.

(a) (b)

Figure 7. Different types of video footage: (a) frame extracted from La Liga game, (b) frame extracted from a third division game.

3.2 Metrics

Bearing in mind that OpenPose detected upper-torso parts in **89.69%** of the given image crop, the following metrics were validated with sensor data:

***LR-Side*:** this metric shows the accuracy of the *LR-side* parameter, which indicates if a player is facing the left or the right side of the field. Considering a sequence of duration T and being i an individual player in a total of NP_t players in frame t, pose orientation α_{i_t} can be obtained as the middle value of the maximum weighted bin in H_{Pi_t}. Having the corresponding ground-truth orientation ω_{i_t} , this metric can be computed as:

$$\mathrm{LR}_{\mathrm{acc}} = \frac{\sum_{t=0}^{T} \sum_{i=0}^{NP_t} LRV_{i_t}}{\sum_{t=0}^{T} NP_t}$$

where

$$LRV_{i_t} = \begin{cases} 1 & \text{if } |\alpha_{i_t} - \omega_{i_t}| < |\alpha_{i_t} + 180 - \omega_{i_t}| \\ 0 & \text{otherwise} \end{cases}$$

LR-Side performance reached **92.43%** accuracy.

Orientation: the comparison between obtained orientation (α_{i_t}) results and ground-truth data (ω_{i_t}) can be displayed by appending all orientation differences in a single vector OD with entries:

$$OD_{i_t} = \min(|\alpha_{i_t} - \omega_{i_t}|, 360 - |\alpha_{i_t} - \omega_{i_t}|)$$

3.3 Parameters Adjustment

By testing all possible weight combinations (in 0.05 intervals), results in Figure 8 indicate the error margin of different tests given by the median, showing the performance of each individual orientation estimation and their best mixture.

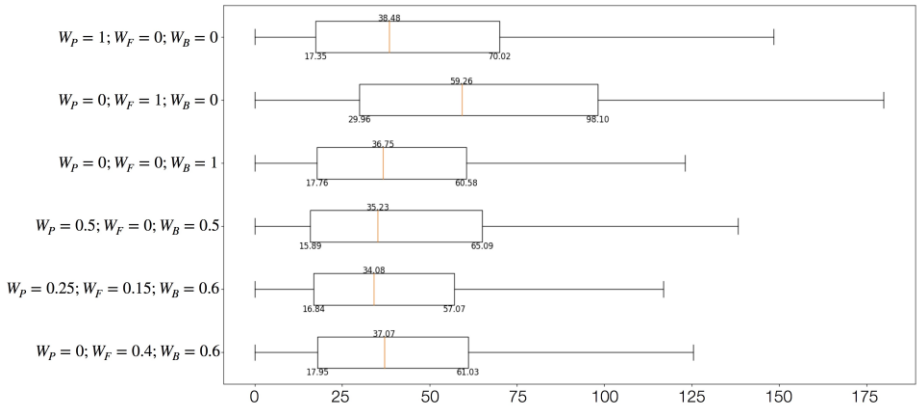

Figure 8. Boxplot results: orientation error quartiles with respect to ground truth. The first 3 rows indicate the performance of pose, field and ball orientation separately; the last three rows indicate the performance of three successful weight combinations (details in the text).

Field orientation is the one producing less accurate orientations, reaching a median absolute error of 59.2567 degrees; although pose orientation performs reasonably better, ball orientation outperforms the rest of individual estimations, both in terms of mean and median absolute errors. Nevertheless, all displayed combination of weights produce better results than any method on their own, keeping ball orientation as the higher-weighted estimation. Anyway, ball orientation alone might fails to extract the appropriate orientation, such as in Figure 9 hence pose and field estimations might help correcting it. The best weight combination has proven to be $W_P = 0.25, W_F = 0.15, W_B = 0.6$, obtaining a median absolute orientation error of **34.075 degrees** and a mean absolute error of **40.007 degrees**. Failure cases include a large number of players looking straight up (90º) and down (270º), as the *LR-side* parameter divides the histogram only in 2 sides, and extreme cases are always challenging. Despite not being validated with La Liga games due to non-existent ground-truth orientation data, better results are supposed to be obtained with better video footage (Figure 7).

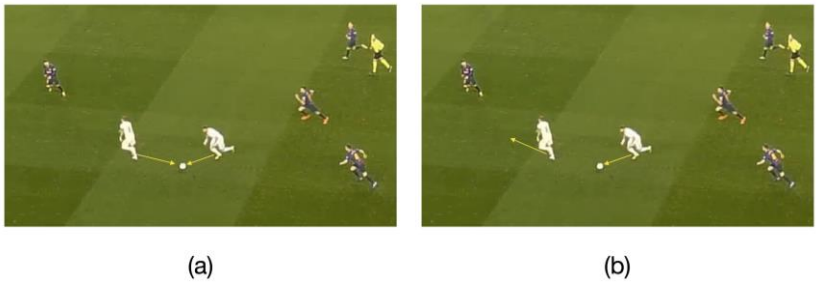

Figure 9. Orientation results with two different weight combinations: (a) only ball orientation, wrong result; (b) pose + field + ball orientation, correct result.

Visual results are displayed in Figure 10.

Figure 10. Visual results, displayed in (left) the original frames, and (right) the 2D field.

4 Practical Applications

Having a time-based set of player orientations, several applications can be found. Among many others:

- Orientation assessment through video. Football analysts spend a lot of time preparing scouting video sessions. By adding orientation information in a visual way such as in Figure 10, they might detect specific situations where orientation is crucial.
- Refining pass probability and pitch-control models. Finding the free man is one of the biggest concerns in offensive football plays. Recent literature has introduced an EPV metric [4] that indicates the conditional probability of scoring/receiving a goal given the ball/player movement and spacing. Logically, players with no defense in open-court maximize this metric; however, the orientation of these players is not being taken into account. Figure 11 shows an example where the best candidate to receive the ball is not facing the passing player.
- Evaluation of on-ball and off-ball direction. Anatomical patterns can easily spotted from pose and orientation data, especially in youth teams. For instance, left-footed players usually start dribbling towards their left side of the court, as their corporal position is not optimal when facing the other side. Orientation metrics could help identifying which players struggle more with left-right deviations.
- Reaction time measurement. Football players must react fast in all scenarios and switch offensive-defensive roles on the fly in short time intervals. By having the players' orientation and eventing data, reaction time can be measured and contextualized.

Figure 11. In the given play, a model to find the best potential passes is used; white circles are printed around those players with better advantage to create an offensive play. However, the passer cannot see one of them due to an inappropriate orientation.

5 Conclusions

In this article, a novel method to compute football players' orientation from a video has been presented. It combines three different orientation estimators: pose, field and ball. Pose orientation is obtained by projecting OpenPose output on a 2D space and computing the normal vector to the projected torso. Field orientation approximates the players' field of view based on their position. Finally, the ball estimation calculates the orientation of all players with respect to the ball weighted by their pairwise distances. Having discretized all orientations into probability vectors, a simple weighting is performed and an individual angle (in degrees) is obtained for each player.

Results have been tested and validated with professional football matches; although the associated video footage was not optimal, results are promising. 92% accuracy is obtained in left-right side orientations, and a median absolute error of 34.07^{o} is achieved. The output of this work is a set of raw time-based numerical orientations that can later be applied in video assessment or refinement of existing models. As future work, besides improving the median absolute error, which could be done by training a deep learning model plus adding eventing data in the weighting step, the generalization to other sports and scenarios will be studied. Furthermore, other applications using pose data could be tested, such as performing team/individual action recognition.

6 Acknowledgements

The authors acknowledge partial support by MICINN/FEDER U project, reference PGC2018-098625-B-I00, H2020-MSCA-RISE-2017 project, reference 777826 NoMADS and F.C. Barcelona's data support.

References

[1] *OpenPose*. https://github.com/CMU-Perceptual-Computing-Lab/openpose. Last accessed: October 13th 2019.

[2] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. *Realtime multi-person 2d pose estimation using part affinity fields.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 7291-7299, 2017.

[3] Francesco Cardinale *et al. Image Super Resolution*. https://github.com/idealo/image-super-resolution, 2018.

[4] Javier Fernández, Luke Bornn, and Dan Cervone. *Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer.* In 13th MIT Sloan Sports Analytics Conference, 2019.

[5] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision.* Cambridge University Press, 2003.

[6] Kevin Palmer and Gerry Geurts. *The evolution of curling analytics.* In 13th MIT Sloan Sports Analytics Conference, 2019.

[7] Varun Ramakrishna, Daniel Munoz, Martial Hebert, James Andrew Bagnell, and Yaser Sheikh. *Pose Machines: Articulated pose estimation via inference machines.* In European Conference on Computer Vision, pages 33-47. Springer, 2014.

[8] Thomas Seidl, Aditya Cherukumudi, Andrew Hartnett, Peter Carr, and Patrick Lucey. *Bhostgusters: Realtime interactive play sketching with synthesized NBA defenses.* MIT Sloan Sports Analytics Conference, 2018.

[9] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. *Convolutional pose machines.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4724-4732, 2016.

[10] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. *Residual dense network for image super-resolution.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2472-2481, 2018.