ALWAYS LOOK ON THE BRIGHT SIDE OF THE FIELD: MERGING POSE AND CONTEXTUAL DATA TO ESTIMATE ORIENTATION OF SOCCER PLAYERS

A. Arbués-Sangüesa¹, A. Martín¹, J. Fernández², C. Rodríguez², G. Haro¹, C. Ballester¹

¹Universitat Pompeu Fabra, ²Futbol Club Barcelona

ABSTRACT

Although orientation has proven to be a key skill of soccer players in order to succeed in a broad spectrum of plays, body orientation is a yet-little-explored area in sports analytics’ research. Despite being an inherently ambiguous concept, player orientation can be defined as the projection (2D) of the normal vector placed in the center of the upper-torso of players (3D). This research presents a novel technique to obtain player orientation from monocular video recordings by mapping pose parts (shoulders and hips) in a 2D field by combining OpenPose with a super-resolution network, and merging the obtained estimation with contextual information (ball position). Results have been validated with players-held EPTS devices, obtaining a median error of 27 degrees/player. Moreover, three novel types of orientation maps are proposed in order to make raw orientation data easy to visualize and understand, thus allowing further analysis at team- or player-level.

Index Terms— Soccer, Orientation, Sports Analytics, Pose, Data Visualization.

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 soccer [8], to the forecasting of future movement locations in basketball [21], tracking data has provided a rich source of information for exploring complex spatio-temporal dynamics in team sports (e.g., [11, 12, 14, 18, 20, 23, 24]). 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. Proper orientation has proven to be crucial for soccer players in order to excel in particular situations, such as receiving or giving ideal passes because of an appropriate field of view, defending two players at a time or finding open-spaces due to a fast reaction. Body orientation is claimed to be more meaningful in the sports’ context than gaze orientation (given by methods such as [16, 10]); nevertheless, only few contributions have been made about body orientation in sports’ challenging scenarios [7, 5]. The main goal of this article is to estimate the body orientation of soccer players from video data, with potential generalization to other sports. By seeking the 2D orientation of the field projection of the normal vector placed in the center of the upper-torso of players, this paper presents a novel technique to extract orientation by merging pose and contextual information. On the one hand, OpenPose [1] is used in combination with a super-resolution network [26] to extract the coordinates of body parts of every single player; projectively mapping key pose parts (in particular, shoulders and hips) in a 2D field-space results in a first orientation estimation; on the other hand, contextual information quantifies the orientation of each player with respect to the ball. The interaction between the ball and the players has been acknowledged as important for action analysis (e.g., [17, 6, 15, 22]). Results have been obtained by validating the output of the presented method with data extracted from players-held EPTS [9] devices: 96.5% accuracy on left-right orientation directions is obtained together with an absolute median error of 27.66 degrees/player; a visual example can be seen in Fig. 1. Besides, as raw orientation data might be difficult to interpret, three different visualization tools are proposed to analyze the orientation information in relation with different events or match context: OrientSonar, Reaction, and On-field maps.

2. PROPOSED METHOD

In this paper, the orientation of a player’s body is defined as the rotation of the player’s upper-torso about the vertical axis, which is assumed to coincide with the angle of the 2D field projection of a 3D 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 Fig. 2. This Section provides a detailed explanation of two different kinds of orientation estimation from...
Fig. 2. Proposed pipeline. On the one hand, pose orientation is found by combining a super-resolution network, OpenPose and 3D vision techniques (plus a coarse validation); on the other hand, ball orientation is also computed.

which the method benefits: (1) pose data and (2) 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 Fig. 3(a). While the orientation value indicates the bin with higher probability, the confidence value is used as a prior to quantify, in an inversely proportional way, 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 estimated angle (maximum of pdf) and its confidence (inverse of the pdf support). A posteriori, a contextual weighting is performed to finally output the orientation of each player.

2.1. Pose Orientation

Estimating orientation from pose data is a key ingredient of our method, 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, the normal vector between these points can be extracted.

**Pose Detection:** having the bounding boxes for all visible players in each frame, the OpenPose library [11] can be used to extract the pose of every single individual (we refer to [19, 25, 2] for details of pose models). Given a soccer frame, the output of the pose estimator is a $25 \times 3$ 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, the average resolution of bounding boxes around players in Full-HD frames is around $15 \times 50$ pixels. Hence, small image crops are not always properly processed by OpenPose, resulting in a null set of 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, 26].

**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 four field corners’ coordinates in the image plane, a homography is computed with the DLT algorithm [13] after establishing the four field-corners correspondences between the frame and a 2D field given by a template image of it. Other homography estimation strategies can be used (e.g., [4]). 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 Fig. 3(b). Basically, the player can be inclined towards the right (0-90°, 270-360°, bins 0-11) or the left (90-270°, bins 12-23) side of the field. 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.

Fig. 4 shows in more detail 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 ($\alpha_{Sh}, \alpha_{Hi}$) and confidences ($C_{Sh}, C_{Hi}$) are obtained, where the suffixes $Sh$ and $Hi$ stand for shoulders and hips, respectively. 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: If $C_{Sh} > C_{Hi}$, $\alpha_{Sh}$ is considered as the pose orientation estimation and $C_{Sh}$ its confi-
Coarse orientation validation: despite the notable performance of OpenPose, image quality problems (e.g. blurry or really small players) are challenging scenarios where estimated players’ pose might be flipped 180°: this is, the right-left shoulders (or hip parts) of the corresponding player are swapped. An inaccurate detection of the pose results in huge errors while estimating the pose angle, as the actual normal vector is the opposite of the predicted one, thus introducing errors that might oscillate between 120° and 180°.

In order to double-check the pose orientation estimation and to ensure that the upper-torso normal vector is computed in the correct direction, a Support Vector Machine model has been trained to classify three types of coarse orientations: front-, side- and back-oriented players (see Fig. 5). Two characteristics are concatenated in the feature vector: color features in the Hue-Saturation-Value color space (histogram of 36-18-18 bins in the respective channel) and geometrical properties (pixel-wise distances between the 4 upper-torso coordinates). Having the position of the upper-torso parts, obtained from pose keypoints, the above-mentioned features are only computed inside the defined trapezoid, hence discarding misleading features such as the color of the field.

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

2.3. Contextual Merging

Once both histograms are obtained, a simple weighting is performed between them, thus merging pose and ball orientations. In particular: \( H_{TOT} = wH_P + (1 - w)H_B \), with \( w \in [0, 1] \). The orientation \( \theta \) of each player is the central value of the bin \( H_{TOT} \) with higher weight.

3. RESULTS

The dataset provided by F.C. Barcelona included video footage (25 fps) of several games from La Liga, tracking data in both frame and field domains, corners positions and contextual information. Moreover, XYZ orientation data were gathered from youth games using EPTS devices [9] for qualitative assessment.

Bearing in mind that OpenPose detected upper-torso parts in 89.69% of the given image crops:

**Coarse orientation validation:** 14,000 players were manually labelled (front, back or side); by randomly splitting it into train and test (80-20), 85.91% accuracy was obtained.
Table 1. MEAE and MDAE given different weights.

| $w$   | $(1 - w)$ | MEAE | MDAE |
|-------|-----------|------|------|
| 0     | 1         | 35.33| 31.59|
| 1     | 0         | 29.98| 27.75|
| 0.3   | 0.7       | 33.77| 29.87|
| 0.7   | 0.3       | 29.78| 27.66|

**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 $i_t$ an individual player in a total of $NP_t$ players in frame $t$, pose orientation $\alpha_{i_t}$, and the corresponding ground-truth orientation $\omega_{i_t}$, this metric can be computed as:

$$LR_{\text{acc}} = \frac{\sum_{t=0}^{T} \sum_{i_{t}=0}^{NP_{t}} LRV_{i_{t}}}{\sum_{t=0}^{T} NP_{t}}$$

where:

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

$LR_{\text{side}}$ performance reached 96.57% accuracy.

**Parameter Adjustment:** by testing all possible weight combinations (in 0.05 intervals), results in Table 1 indicate the error margin of different tests, showing the performance of each individual orientation estimation and their best mixture. As it can be observed, ball orientation produces the less accurate predictions; pose orientation outperforms this prediction by a notable margin. These individual results prove that pose orientation needs to be heavily weighted while merging both estimations: by setting $w$ to 0.7, the mean absolute angle error (MEAE) is reduced to 29.78° and the median absolute angle error (MDAE) to 27.66°.

### 5. CONCLUSIONS

In this article, a novel technique to compute soccer players’ orientation from a video has been presented. The method combines two different orientation estimators: pose and ball. While pose orientation is obtained by projecting OpenPose output on a 2D space and computing the normal vector to the projected torso, ball estimation calculates the orientation of all players with respect to the ball. Results have been tested and validated with professional soccer matches: 96.6% accuracy is obtained in left-right side orientations, and a median absolute error of 27.66° is achieved. As future work, besides improving the MEAE, which could be done by training an end-to-end deep learning model, the generalization to other sports and scenarios will be studied. Furthermore, other applications using pose data could be tested (i.e., team/individual action recognition). Finally, new types of visualizations could be designed together with complex game phases by including more information in the dataset.
6. REFERENCES

[1] OpenPose. https://github.com/CMU-Perceptual-Computing-Lab/openpose.

[2] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Real-time multi-person 2d pose estimation using part affinity fields. In IEEE Conference on Computer Vision and Pattern Recognition, pages 7291–7299, 2017.

[3] F. Cardinale. Isr. https://github.com/idealo/image-super-resolution, 2018.

[4] J. Chen and J. J. Little. Sports camera calibration via synthetic data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 0–0, 2019.

[5] M. Fastovets, J.-Y. Guillemaut, and A. Hilton. Athlete pose estimation from monocular tv sports footage. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 1048–1054, 2013.

[6] P. Felsen, P. Agrawal, and J. Malik. What will happen next? forecasting player moves in sports videos. In IEEE ICCV, pages 3342–3351, 2017.

[7] P. Felsen and P. Lucey. Body shots: Analyzing shooting styles in the NBA using body pose. In MIT Sloan, Sports Analytics Conference, 2017.

[8] J. Fernández, L. Bornn, and D. Cervone. Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. In 13th MIT Sloan Sports Analytics Conference, 2019.

[9] FIFA. EPTS Electronic Performance and Tracking Systems. https://football-technology.fifa.com/en/media-tiles/epts/.

[10] T. Fischer, H. Jin Chang, and Y. Demiris. Rt-gene: Real-time eye gaze estimation in natural environments. In European Conference on Computer Vision (ECCV), pages 334–352, 2018.

[11] J. Gao, T. Zhang, and C. Xu. Graph convolutional tracking. In IEEE Conference on Computer Vision and Pattern Recognition, pages 4649–4659, 2019.

[12] R. Girdhar, G. Gkioxari, L. Torresani, M. Paluri, and D. Tran. Detect-and-track: Efficient pose estimation in videos. In IEEE Conference on Computer Vision and Pattern Recognition, pages 350–359, 2018.

[13] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003.

[14] S. Jin, W. Liu, W. Ouyang, and C. Qian. Multi-person articulated tracking with spatial and temporal embeddings. In IEEE Conference on Computer Vision and Pattern Recognition, pages 5664–5673, 2019.

[15] P. R. Kamble, A. G. Keskar, and K. M. Bhurchandi. Ball tracking in sports: a survey. Artificial Intelligence Review, 52(3):1655–1705, 2019.

[16] P. Kellnhofer, A. Recasens, S. Stent, W. Matusik, and A. Torralba. Gaze360: Physically unconstrained gaze estimation in the wild. In IEEE International Conference on Computer Vision (ICCV), 2019.

[17] A. Maksai, X. Wang, and P. Fua. What players do with the ball: A physically constrained interaction modeling. In IEEE conference on computer vision and pattern recognition, pages 972–981, 2016.

[18] M. Manafifard, H. Ebadi, and H. A. Moghaddam. A survey on player tracking in soccer videos. Computer Vision and Image Understanding, 159:19–46, 2017.

[19] V. Ramakrishna, D. Munoz, M. Hebert, J. A. Bagnell, and Y. Sheikh. Pose machines: Articulated pose estimation via inference machines. In European Conference on Computer Vision, pages 33–47. Springer, 2014.

[20] N. Ran, L. Kong, Y. Wang, and Q. Liu. A robust multi-athlete tracking algorithm by exploiting discriminant features and long-term dependencies. In International Conference on Multimedia Modeling, pages 411–423. Springer, 2019.

[21] T. Seidl, A. Cherukumudi, A. Hartnett, P. Carr, and P. Lucey. Bhostgusters: Realtime interactive play sketching with synthesized NBA defenses. In MIT Sloan Sports Analytics Conference, 2018.

[22] G. Thomas, R. Gade, T. B. Moeslund, P. Carr, and A. Hilton. Computer vision for sports: Current applications and research topics. Computer Vision and Image Understanding, 159:3–18, 2017.

[23] N. Wang, Y. Song, C. Ma, W. Zhou, W. Liu, and H. Li. Unsupervised deep tracking. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1308–1317, 2019.

[24] X. Wang, A. Jabri, and A. A. Efros. Learning correspondence from the cycle-consistency of time. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2566–2576, 2019.

[25] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. In IEEE Conference on Computer Vision and Pattern Recognition, pages 4724–4732, 2016.

[26] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2472–2481, 2018.