A Method for Player Importance Prediction from Player Network Using Gaze Position Estimated by LSTM

Genki Suzuki (student member)†, Sho Takahashi (member)††, Takahiro Ogawa (member)†††, Miki Haseyama (fellow)†††

Abstract  A novel method for player importance prediction from a player network using gaze positions estimated by Long Short-Term Memory (LSTM) in soccer videos is presented in this paper. By newly using an estimation model of gaze positions trained by gaze tracking data of experienced persons, it is expected that the importance of each player can be predicted. First, we generate a player network by utilizing the estimated gaze positions and first-arrival regions representing players’ connections, e.g., passes between players. The gaze positions are estimated by LSTM that is newly trained from the gaze tracking data of experienced persons. Second, the proposed method predicts the importance of each player by applying the Hypertext Induced Topic Selection (HITS) algorithm to the constructed network. Consequently, prediction of the importance of each player based on soccer tactic knowledge of experienced persons can be realized without constantly obtaining gaze tracking data.

Key words: Sports video analysis, tactical analysis, first-arrival region, link analysis, gaze tracking data, long short-term memory.

1. Introduction

In recent years, with increasing popularization of sports video distribution services such as DAZN* and MyCujoo**, the number of soccer videos has been increasing [1]. However, it is difficult for various viewers to fully understand the soccer videos since viewers have different levels of knowledge regarding to soccer tactics and rules of soccer. Therefore, it is effective for a semantic understanding of soccer videos to present information obtained from the analysis of tactics such as estimation of pass courses and off-the-ball movements of players.

Many researchers have proposed methods of tactical analysis from soccer videos [2–11]. Analysis is conducted by these methods for obtaining semantic information such as information on movements of players [2–4], quantification of possibilities of shots and passes [5], estimation of pass courses [6], estimation of team formations [7–9] and importance prediction of each player [10,11]. Such information is useful for understanding team tactics, and we have focused on predicting the importance of each player who has an important role in soccer games [10,11].

In [10], the importance of each player was predicted by analyzing a player network generated on the basis of first-arrival regions [12], which represent neighboring players’ connections by short passes between nearby players. Furthermore, in [11], it became possible to generate a new network by introducing gaze tracking data of experienced persons. This enables consideration of new players’ connections such as long and through passes between a ball holder and ball receivers in distant places. The gaze tracking data of experienced persons include tactic knowledge, and these data enable consideration of movements and positions of players and spaces [13]. Then effective prediction of the importance of each player in various attack scenes has been achieved. However, in the previous method [11], in order to predict the importance of each player, it is necessary to obtain gaze tracking data of experienced persons for all of the target soccer videos. Generally, it is difficult and non-realistic to obtain gaze tracking data from experienced persons for all soccer videos.
solve the above problem, it is necessary to construct a new method in which automatic estimation of gaze positions is introduced into the previous method [11].

In recent years, improvement of the accuracy of sports video analysis using deep learning, which is one type of machine learning, has been reported [14–16]. In addition, preparation of a large amount of training data such as player and ball position data is needed for constructing a deep learning network, and these data can often be obtained by systems such as the TRACAB image tracking system*, which is used in the FIFA World Cup and by many professional leagues [17]. Therefore, it is expected that the importance of each player can be predicted by using a deep learning model trained by the player and ball positions and gaze tracking data.

A new method for predicting the player importance based on a player network that includes a gaze position estimation model is presented in this paper. The proposed method is composed of the following two stages: 1) generation of a player network using the estimated gaze positions and 2) player importance prediction based on the HITS algorithm. Figure 1 shows an overview of our method. In the first stage, the proposed method generates the player network based on first-arrival regions and estimated gaze positions. Specifically, the edge’s strength of the player network is calculated on the basis of the first-arrival regions for representing the players’ connections. In this calculation, gaze positions estimated by a Long Short-Term Memory (LSTM) [18] model are also considered. The LSTM model is a deep learning model for recognizing patterns of sequential data. Therefore, it enables training of gaze trends of experienced persons. As a result, the generation of a player network based on the players’ relationships and tactic knowledge of experienced persons is realized. In the second stage, the proposed method predicts the importance of each player by applying the HITS algorithm [19] utilized for link analysis of Web pages to the generated player network. Consequently, the new framework in our method contributes to improvement in the performance of prediction of the importance of each player without constantly obtaining gaze tracking data from experienced persons.

2. Player Importance Prediction

In this section, we explain the proposed method for player importance prediction. This importance is the degree of connections with other players by passes. By calculating the importance, it is possible to extract players who play important roles in terms of soccer tactics. Specifically, a player with high importance is more likely to receive a pass than other players. In our method, the importance is calculated by analyzing the network representing the players’ connections by using the HITS algorithm. The network is generated on the basis of the first-arrival regions and estimated gaze positions. The details are shown below.

2.1 Generation of the Player Network

In this section, we explain the generation of the player network. First, we define the player network by using the adjacency matrix $L_t = [L_t^{mn}]$ of the players where the players of the attacking team are shown as $p_m^t$ and $p_n^t$ ($m, n = 1, 2, \ldots, 11; m \neq n; t = 1, 2, \ldots, T; T$ being the total number of frames). The matrix $L_t$ is defined

---

* https://chyronhego.com/products/sports-tracking/tracab-optical-tracking/
by the edge’s strength based on the first-arrival regions and estimated gaze positions as follows:

\[
|L_t^{mn}| = \begin{cases} 
\alpha \frac{d t_{mn}^t}{dt} + (1 - \alpha) w_{t}^{mn} & (m \neq n) \\
0 & \text{(otherwise)}, \end{cases}
\] (1)

where \(l_{t}^{mn}\) is the length of the boundary line of the first-arrival region between players \(p_t^m\) and \(p_t^n\), and \(d_{t}^{mn}\) is the distance between these players. Furthermore, \(w_{t}^{mn}\) is the edge’s strength based on the gaze, and \(\alpha\) is an adjustment weight \((0 \leq \alpha \leq 1)\). In Eq. (1), the first term gives the edge’s strength based on the first-arrival regions, and the second term gives the edge’s strength based on the estimated gaze positions that reflect tacit knowledge.

(1) Calculation of the Edge’s Strength Based on First-arrival Regions

We explain the calculation of the edge’s strength of the player network based on the first-arrival regions. The first-arrival region is defined as the region in which each player first arrives in the pitch [12]. It is calculated by using the minimum arrival time to arbitrary points on the field based on the velocity and position of each player. The movement model of the player \(p_t^m\) for generation of the first-arrival regions is defined as follows:

\[
m_{\text{player}} \frac{dv^m}{dt} = F - kv^m, \tag{2}
\]

where \(m_{\text{player}}\) is the mass of the player, \(\frac{dv^m}{dt}\) is differentiation in time \(s\) of velocity \(v^m\) of the player, \(F\) is the maximum vector of propulsion of the player, and \(k\) is a resistance coefficient. Furthermore, \(F = Fe\), where \(F\) is the maximum propulsion and \(e\) is a unit vector. Note that the second term on the right side is a resistance force to the movement of the player based on [20] to ensure that the speed does not become infinite with time. Next, Eq. (2) is rewritten as follows:

\[
x^m - x^m_t = V_{\text{max}} \left( s - \frac{1 - \epsilon^{-\beta s}}{\beta} \right) + \frac{1 - \epsilon^{-\beta s}}{\beta} v^m_t, \tag{3}
\]

where the above equation is a solution to the differential equation of Eq. (2), \(x^m\) is a position of the player, \(x^m_t\) is the current position of \(p_t^m\), \(V_{\text{max}}\) is maximum velocity of the player, and \(v^m_t\) is the current velocity of \(p_t^m\). Furthermore, \(\beta\) denotes the magnitude of the resistance. From Eq. (2), since \(v^m < \frac{F}{k}\) is satisfied, Eq. (3) indicates that a player with \(v^m_t\) at \(x^m_t\) can reach the point \(x^m\) in the following time length \(s\). Finally, the first-arrival region \(D\) where the player \(p_t^m\) can arrive earlier than any other players is calculated on the basis of Eq. (3) as follows:

\[
D(p_t^m) = \{ x^m_{\text{pnt}} \in R^2 | t_s(x^m_{\text{pnt}}, p_t^m) \leq t_s(x^m_{\text{pnt}}, p_n^m), \}
\]

\[
|L_t^{mn}| \leq \{ m \neq n \}, \tag{4}
\]

where \(t_s(x^m_{\text{pnt}}, p_t^m)\) is the shortest time to arrival at arbitrary point \(x^m_{\text{pnt}}\) of player \(p_t^m\). Generally, the possibility of a successful pass becomes high when the first-arrival regions of two players belonging to the same team are adjacent. Therefore, when the boundary line of the first-arrival region is longer and the distance between the two players is smaller, the possibility of a successful pass becomes higher. By calculating the edge’s strength of the player network based on the first-arrival region that represents players’ connections, the proposed method generates the player network. The proposed method calculates \(l_{t}^{mn}\) and \(d_{t}^{mn}\) by using the first-arrival region. Figure 2 shows an example of the first-arrival region, \(l_{t}^{mn}\) and \(d_{t}^{mn}\).

In this paper, we focus on the task of estimating important players who use the tactics that will immediately lead to the goal in a few steps. Therefore, we do not consider back-passes when constructing the player network based on the first-arrival regions. Thus, when player \(p_t^m\) exists behind player \(p_t^n\), the possibility of a pass between these two players becomes low. Hence, \(|L_t^{mn}|\) is set to 0 in this case. It should be noted that back-passes by crosses can be considered in the proposed method since the relationship between the ball holder and candidates of the ball receiver tends to be represented by the player network based on estimated gaze positions. Its details are shown below.

(2) Calculation of the Edge’s Strength Based on Estimated Gaze Positions

Generally, experienced persons focus on soccer tactics while watching soccer videos. Specifically, experienced
soccer players look at the movements and positions of players, which are closely related to the tactics, and spaces more often than do inexperienced persons [13]. Therefore, the gaze tracking data of experienced persons are effective for determining the relationship between the ball holder and the ball receiver. In the proposed method, we train a deep learning-based estimation model of the gaze position by using the gaze tracking data as ground truth and features extracted from player and ball position data. These data and features are shown in Table 1. Consequently, it is possible to obtain gaze positions with the extracted features without gaze tracking data. After that, we calculate the edge’s strength between the ball holder and the ball receiver in the player network by using the gaze position that is obtained by inputting the extracted features to the estimation model. As a result, generation of a network that includes tactic knowledge of experienced persons is realized.

In the rest of this subsection, the feature extraction is first described in detail. Then the construction of the LSTM-based model for estimating the gaze position is described. Finally, we explain the generation of the player network based on the gaze position.

### A. Extraction of Features and Gaze Positions

First, we extract features \( u_t \) from the player and ball positions as shown in Table 1. Formation features can effectively express the characteristics of team tactics based on player positions which are the distribution of players on the soccer field and behaviors of individual players [8]. In addition, we train the model by using the player and ball positions with the formation features. Next, we obtain the gaze tracking data \( d_t \) from \( S \) experienced persons. In the proposed method, the model for estimating the gaze position is trained by using these features and the ground truth data. Note that the gaze position of experienced persons is translated from the screen coordinate system to the same coordinate system as that of the player and ball position data by utilizing the previously reported method [21].

### B. Construction of Gaze Position Estimation Model Based on LSTM

In recent years, recurrent neural networks (RNNs) have been shown to have remarkable performance in many sequential learning problems such as video recognition, natural language processing and speech recognition [22–24]. An RNN is a deep learning model that recognizes patterns of sequential data and has a cycle inside the network. Since the model of an RNN includes temporarily memorized information, it can train sequential data. Since gaze tracking data are sequential data, the estimation model of gaze positions in soccer videos can be constructed by using an RNN. Specifically, we estimate the gaze position based on LSTM [18] which is one of the RNN since it is better at finding and exploiting long-range context in gaze tracking data with memory cells to store information. Figure 3 shows the structure of the LSTM. The LSTM consists of an input gate \( i_t \), an output gate \( o_t \), a forget gate \( f_t \), and a cell \( c_t \) as follows:

\[
\begin{align*}
    i_t &= \sigma(Z_{ui}u_t + Z_{hi}h_{t-1} + Z_{ci}c_{t-1} + b_i) \\
    f_t &= \sigma(Z_{uf}u_t + Z_{hf}h_{t-1} + Z_{cf}c_{t-1} + b_f) \\
    c_t &= f_t c_{t-1} + i_t \tanh(Z_{wc}u_t + Z_{hc}h_{t-1} + b_c) \\
    o_t &= \sigma(Z_{wo}u_t + Z_{ho}h_{t-1} + Z_{co}c_t + b_o) \\
    h_t &= o_t \tanh(c_t),
\end{align*}
\]

where \( u_t \) and \( h_t \) are the input vector and output vector, respectively. Furthermore, \( i_t, f_t, c_t \) and \( o_t \) are the outputs of each gate. \( Z_{\bullet \bullet} \) shows the weight matrix between each gate. For example, \( Z_{ui} \) shows the weight matrix between the input \( u_t \) and the input gate \( i_t \). In Eqs. (5) - (9), \( b_\bullet \) is a bias, and \( \sigma(\cdot) \) in Eqs. (5), (6) and (8) is a sigmoid function. In the proposed method, the model for estimating the gaze position is constructed by two-dimensional regression using the features and the gaze tracking data. Then we set a cost function \( C \) based on the mean square error (MSE) [25] shown as follows:

\[
    C = \frac{1}{T} \sum_{t=1}^{T} ||h_t - d_t||^2. \tag{10}
\]

In the proposed method, the weight between each gate is trained on the basis of Backpropagation Through Time [26]. In this way, we can perform training of the

| Features : \( u_t \) | Dim. |
|----------------------|-----|
| Players positions    | 44  |
| Ball position        | 2   |
| Formation features [8]| 111 |

| Ground-truth : \( d_t \) |
|--------------------------|
| Gaze position from gaze tracking data | 2 |
LSTM-based estimation of gaze positions.

In a test phase, by inputting a new feature vector $u_t$ to the above model, estimation results $h_t$ of the gaze position corresponding to the target frame are obtained in Eq. (9) from the output layer of the LSTM. Furthermore, to correct their instantaneous errors, a voting scheme is performed for the results of the estimated gaze positions. Specifically, it is performed for $M$-frames before and after the target frame based on [27]. Consequently, our method obtains the gaze position by selecting the most voted positions.

C. Generation of a Player Network Based on Gaze Position

Calculation of the edge’s strength $w_{mn}^t$ utilizing the estimated gaze position is explained. It has been reported that there is a difference between the visual ability of experienced persons and that of inexperienced persons [28]. Specifically, it has been reported that experienced persons have a wider range of perceived movements in peripheral vision than do inexperienced persons. In the previous method [11], a link to only one player is generated on the basis of the distance between the gaze position and each player. For this reason, it is difficult to generate a link for a target player by the previous method when there are multiple players in the gazed area. In order to solve this problem, we modified the link generation method. Specifically, in the proposed method, some links for multiple players are generated on the basis of the distance between the gaze position and each player who is located around the gaze position. By utilizing the gaze position $h_t$ that is predicted on the basis of the previous $T'$ frames, the weight $w_{mn}^t$ is calculated as follows:

$$w_{mn}^t = 1 \frac{1}{T'} \sum_{j=0}^{T'} z_{mn}^{t-j},$$ (11)

$$z_{mn}^{t-j} = \gamma_{mn}^{t-j} \frac{\text{dis}(h_{t-j}, x_{n}^{t-j})}{\text{dis}(h_{t-j}, x_{n}^{t-j}) \leq R)},$$ (12)

where $\text{dis}(h_{t-j}, x_{n}^{t-j})$ is the Euclidean distance between the gaze position $h_{t-j}$ and the position $x_{n}^{t-j}$ of player $p_n^t$, and $\gamma_{mn}^{t-j}$ is a normalization parameter to bring $\text{dis}(h_{t-j}, x_{n}^{t-j})$ into the range $[0, 1]$. $z_{mn}^{t-j}$ indicates that the edge’s strength of player $p_n^t$ who is located in the circular region with the radius $R$ centered on $h_{t-j}$ and decreases with the distance between $h_{t-j}$ and $x_{n}^{t-j}$. Figure 4 shows an example of generating the player network by using the first-arrival region and the gaze position. As a result, the surrounding players can be weighted on the basis of the peripheral vision.

2.2 Player Importance Prediction based on the HITS Algorithm

In this section, we explain the prediction of the importance of each player in the player network based on the HITS algorithm. The HITS algorithm extracts the authority and hub scores of each node included in the target networks based on link analysis [19]. A node that has a high authority score is linked from many nodes. On the other hand, a node that has a high hub score links to many authority nodes. The HITS algorithm can find important nodes by using these scores. Then the importance of each player is predicted by utilizing the authority score based on the HITS algorithm. In the proposed method, the authority score is defined as
the importance of each player who is a candidate for the ball receiver. By applying the HITS algorithm to the player network $L_t$ obtained as described in section 2.1, we can obtain the importance of each player. Specifically, the importance $auth_t^c(m)$ ($c = 1, 2, \ldots, C; C$ being the total number of iterations) of player $p_t^m$ is defined as follows:

$$auth_t^c(m) = \varepsilon_t \sum_{n=1}^{11} L_t^{nm} \cdot hub_t^{c-1}(n), \quad (13)$$

$$hub_t^c(m) = \eta_t \sum_{n=1}^{11} L_t^{mn} \cdot auth_t^c(n), \quad (14)$$

where $hub_t^c(m)$ is a hub score, and $\varepsilon_t$ and $\eta_t$ are normalized constants. Note that both the authority score $auth_t^0(m)$ and the hub score $hub_t^1(m)$ are initially set to one. In this method, by iteratively updating the authority and hub scores for each player, their converged results $auth_t^C(m)$ are obtained for all of the players.

Consequently, the proposed method can effectively predict the importance of each player by applying the HITS algorithm to the player network generated from the first-arrival regions and the estimated gaze positions.

3. Experimental Results

The effectiveness of the proposed method was shown by using actual soccer videos. The experimental conditions are explained in 3.1 and results of the experiments are shown in 3.2.

3.1 Experimental Conditions

Soccer videos and tracking data for players and a ball (3,322 seconds, 30 fps) were used in the experiment. Player position data of the video clip were obtained using the TRACAB optical tracking system*. By applying Tobii Eye Tracker 4C**, gaze tracking data were obtained from 3 subjects ($S = 3$) as experienced persons who had 10 years, 7 years and 3 years of experience as attacking position players. In order to make the differences between gaze positions of the subjects small even though the number of subjects was small, we set a simple task by instructing the subjects to watch the video clip from the viewpoint of the attacking team. Specifically, we asked the subjects to watch soccer videos with seeking the next ball receivers who will contribute to the goal of the attacking team. The gaze position estimation model was constructed by training 2,943 seconds of soccer videos with a single NVIDIA GeForce GTX 2080 Ti GPU. The test data were 57 pass scenes ($D = 57$) extracted from soccer videos. The average time of the scenes was 8 seconds and the total time was 379 seconds. The ground truth was the player who received the ball. We used Recall@$k$ for a quantitative evaluation criterion defined as

$$\text{Recall}@k = \frac{r_k}{D} \quad (k = 1, 2, \ldots, K), \quad (15)$$

where $r_k$ is the number of scenes where the player who actually received the ball was included in the top $k$ players with high importance. $K$ indicates the number of players to be analyzed. In this experiment, since attack scenes were targeted, $K$ was set to 5, which was half of the total number of attack players except for the goalkeeper.

In order to confirm the effectiveness of the proposed method, the following seven methods were adopted for the ideal and comparative methods. We denote the proposed method as PM in this experiment.

**Ideal method (IM)** [11]

This is an ideal version of our method that utilizes the player network generated on the basis of the first-arrival regions and gaze tracking data of experienced persons. The player network is generated by using the actual gaze tracking data of experienced persons. By comparing the level of accuracy of results obtained by using the IM and PM, we verify the reproducibility of the gaze position estimation model.

**Comparative method1 (CM1)** [10]

This is a method that utilizes a player network generated on the basis of only the first-arrival regions. We verify the effectiveness of the use of the gaze position estimated on the basis of the LSTM model.

**Comparative method2 (CM2)**

This is a method that utilizes a player network generated on the basis of the first-arrival regions and the gaze positions based on LSTM that is trained by using player and ball position data without using formation features. We verify the effectiveness of the use of formation features.

**Comparative method3 (CM3)**

This is a method in which LSTM of the proposed method is replaced by a multivariate regression model [29] using Tensorflow [30].

---

* https://chyronhego.com/products/sports-tracking/tracab-optical-tracking/

** https://tobiigaming.com/eye-tracker-4c/
verify the effectiveness of the use of LSTM.

Comparative method4 (CM4)
This is a same method which removes the M-frame voting scheme from PM. We verify the effectiveness of the use of the M-frame voting scheme in the gaze position estimation.

Comparative method5 (CM5)
This is a method that selects the player closest to the actual gaze positions as a player with high importance.

Comparative method6 (CM6)
This is a method that selects the player closest to the gaze positions estimated by LSTM as a player with high importance.

We used CM5 and CM6 to verify the effectiveness of the use of the player network generated on the basis of the first-arrival regions and the HITS algorithm. Details of the parameters used in the proposed method are shown in Table 2. The parameters in each method were determined in such a way that the performance of each method became the highest.

3.2 Performance Evaluation
Recall@k values of all methods are shown in Table 3. From the obtained results, since Recall of PM is closest to IM, it was confirmed that PM outperforms the comparative methods. Specifically, by comparing the results of PM with those of CM1, the effectiveness of the use of the gaze position estimated on the basis of the LSTM model was confirmed. Next, by comparing the results of PM with those of CM2, the effectiveness of the use of formation features was confirmed. Furthermore, since the performance of PM is higher than that of CM3, it was confirmed that the introduction of LSTM is effective. Moreover, by comparing the results of PM with those of CM4, we confirmed that the M-frame voting scheme based on [27] is effective for improving the performance of the proposed method. Finally, we confirmed the effectiveness of the player network collaboratively using the gaze positions and the first-arrival regions to predict the player’s importance by the HITS algorithm since the performances of IM and PM are higher than those of CM5 and CM6, respectively.

Next, in order to verify the importance of α, new experimental results of the proposed method obtained by changing α to $0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1$ are shown in Fig. 5. From the obtained results, we confirmed that the proposed method outputs the best performance when $\alpha = \frac{1}{3}$.

Therefore, we confirmed that it is important to generate the weighted player network based on the estimated gaze positions.

In order to discuss the effects of “the estimation performance of the estimated gaze positions” on “the performance of our player importance prediction”, we focus on the difference in the numbers of video clips for which the player’s importance prediction succeeded and failed in terms of the gaze estimation performance. Specifically, Table 4 shows the relationship between “the difference between the estimated and actual gaze positions” and “success or failure in important player prediction”. From the obtained results, we can see that when the estimated gaze positions are not accurate, the number of failed video clips increases. On the other hand, the number of successful video clips increases when the gaze positions are correctly estimated.

The results of prediction of player importance obtained by using the methods are shown in Fig. 6. In this figure, (a) is an input frame and (b)-(g) are the results of prediction obtained by using the methods. In the actual scene, player $p_1^{11}$ received the ball from player...
$p_5^8$ and made a goal. From the obtained results, we confirmed that the proposed method predicted the high importance of player $p_1^{11}$ as did IM. Specifically, since the link to player $p_1^{11}$ was newly generated by the estimated gaze positions, it became possible to predict the player importance. Therefore, the novelties of the proposed method strongly contribute to the successful prediction of the player importance. However, the difference in the average Recall values between the proposed method and IM is about 12%. Therefore, in order to realize more accurate prediction of the player’s importance by improving the gaze prediction performance, it is necessary to train LSTM by using gaze tracking data obtained from more subjects. This is the future work in our study.

4. Conclusions

A new method for predicting the importance of each player from a player network using gaze position estimated by LSTM in soccer videos is presented in this paper. The proposed method consists of the following two stages. In the first stage, the proposed method generates the player network by utilizing the first-arrival regions and the gaze positions. The gaze position is estimated by LSTM that is trained from the gaze tracking data of experienced persons. Finally, the importance of each player is predicted by applying the HITS algorithm to the obtained network. The proposed method realizes accurate prediction of player importance without constantly obtaining gaze tracking data from experienced persons. Experimental results showed that the proposed method outperforms comparative methods, and the novel approach in our method contributes to the performance improvement. In our future work, we will try to construct a method for prediction of the importance of each player to improve its accuracy.

Acknowledgment

This work was partly supported by JSPS KAKENHI Grant Numbers JP17H01744, JP17K00148.

References

1) H. C. Shih, “A survey of content-aware video analysis for sports,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 5, pp. 1212–1231 (2018)
2) C. Perin, R. Vuillemot, and J. D. Fekete, “SoccerStories: a kickoff for visual soccer analysis,” IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, pp. 2506–2515 (2013)
3) M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, “Action classification in soccer videos with long short-term memory recurrent neural networks,” Artificial Neural Networks, pp. 154–159 (2010)
4) J. Peri, A. Grunz, and D. Mennert, “Tactics analysis in soccer—an advanced approach,” International Journal of Computer Science in Sport, vol. 12, no. 1, pp. 33–44 (2013)
5) Y. Wu, X. Xie, J. Wang, D. Deng, H. Liang, H. Zhang, S. Cheng, and W. Chen, “Forvizor: visualizing spatio-temporal team formations in soccer,” IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, pp. 65–75 (2019)
6) S. Takahashi and M. Haeyama, “Active grid-based pass region estimation from multiple frames of broadcast soccer videos,” ITE Transactions on Media Technology and Applications, vol. 1, no. 3, pp. 220–225 (2013)
7) J. Fernández, L. Bornm, and D. Cervone, “Decomposing the immeasurable sport: a deep learning expected possession value framework for soccer,” in Proceedings of 19th Annual MIT Sloan Sports Analytics Conference (2019)
8) S. Ohnuki, S. Takahashi, T. Ogawa, and M. Haseyama, “Soccer video segmentation based on team tactics estimation method,” in *Proceedings of International Workshop Advanced Image Technology*, pp. 692–695 (2013).

9) G. Suzuki, S. Takahashi, T. Ogawa, and M. Haseyama, “Decision level fusion-based team tactics estimation in soccer videos,” in *Proceedings of IEEE 5th Global Conference on Consumer Electronics*, pp. 58–59 (2016).

10) S. Takahashi and M. Haseyama, “A method of important player extraction based on link analysis in soccer videos,” *ITE Transactions on Media Technology and Applications*, vol. 5, no. 2, pp. 42–48 (2017).

11) G. Suzuki, S. Takahashi, T. Ogawa, and M. Haseyama, “A method for predicting importance of attack players based on multiple gaze tracking data in soccer videos,” in *Proceedings of IEEE 5th Global Conference on Consumer Electronics*, pp. 520–521 (2019).

12) T. Taki and J. Hasegawa, “A feature for group behavior analysis in team sports and its application,” *IEEE Transactions on Information & Systems*, vol. 81, no. 8, pp. 1802–1811 (1998).

13) A. M. Williams, K. Davids, L. Burwitz, and J. G. Williams, “Visual search strategies in experienced and inexperienced soccer players,” *Research Quarterly for Exercise and Sport*, vol. 65, no. 2, pp. 127–35 (1949).

14) K. C. Wang, and R. Zemel, “Classifying NBA offensive plays using neural networks,” in *Proceedings of MIT Sloan Sports Analytics Conference*, pp. 1–9 (2016).

15) T. Tsunoda, Y. Komori, M. Matsugu, and T. Harada, “Football action recognition using hierarchical LSTM,” in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition Workshops*, pp. 99–107 (2017).

16) R. Shah and R. Romijnders, “Applying deep learning to basketball trajectories,” *arXiv preprint arXiv:1608.03793* (2016).

17) C. Kenneth and R. A. Daniel, “The application of sports technology and sports data for commercial purposes,” *The Use of Technology in Sport* (2018).

18) W. Zaremba, I. Sutskever, and O. Vinyals, “Recurrent neural network regularization,” *arXiv preprint arXiv:1409.3239* (2014).

19) J. M. Kleinberg, “Authoritative sources in a hyperlinked environment,” *Journal of the ACM*, vol. 46, no. 5, pp. 604–632 (1999).

20) A. Fujimura and K. Sugihara, “Quantitative evaluation of sport teamwork using generalized Voronoi diagrams,” *IEICE Transactions on Information & Systems*, vol. 38-D-II, no. 3, pp. 818–828 (2004).

21) T. Watanabe, M. Haseyama, and H. Kitajima, “A soccer field tracking method with wire frame model from tv images,” in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 1633–1636 (2004).

22) J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, “Long-term recurrent convolutional networks for visual recognition and description,” in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 2625–2634 (2015).

23) G. Saon, H. K. J. Kuo, S. Rennie, and M. Picheny, “The IBM 2015 English conversational telephone speech recognition system,” *arXiv preprint arXiv:1505.05899* (2015).

24) R. Socher, A. Perelygin, J. Wu, J. Chuang, C.D. Manning, A. Ng, and C. Potts, “Recursive deep models for semantic compositionality over a sentiment Treebank,” in *Proceedings IEEE International Conference on Empirical Methods in Natural Language Processing*, pp. 1631–1642 (2013).

25) E. L. Lehmann and G. Casella, “Theory of point estimation,” *Springer Science & Business Media*, vol. 18, no. 5, pp. 602–610 (2006).

26) A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional lstm and other neural network architectures,” *Neural Networks*, vol. 18, no. 5, pp. 602–610 (2005).

27) G. Suzuki, S. Takahashi, T. Ogawa, and M. Haseyama, “Team tactics estimation in soccer videos based on a deep extreme learning machine and characteristics of the tactics,” *IEEE Access*, vol. 7, pp. 153238–153248 (2019).

28) C. D. Stine and M. R. Arterburn, “Vision and sports: a review of the literature,” *Journal of the American Optometric Association*, vol. 53, no. 8, pp. 627–633 (1982).

29) G. Chamberlain, “Multivariate regression models for panel data,” *Journal of Econometrics*, vol. 18, no. 1, pp. 5–46 (1982).

30) M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, and others, “TensorFlow: Large-scale machine learning on heterogeneous systems,” *Software available from tensorflow.org*, vol. 1, no. 2 (2015).
Fig. 6 Results of importance prediction obtained by the proposed method and the comparative methods. Each figure in (b)-(g) shows the player network generated by the first-arrival region as red arrows and the player network generated by gaze position as orange arrows. The players indicated by black circles are predicted to be players with the highest importance. The purple regions in (b), (c) and (e)-(g) are shown by multiple circular regions based on Eq. (11) with the radius $R$ centered on the actual gaze positions or the estimated gaze positions.