Team Tactics Estimation in Soccer Videos Based on a Deep Extreme Learning Machine and Characteristics of the Tactics

著者	Suzuki Genki, Takahashi Sho, Ogawa Takahiro, Haseyama Miki
journal or publication title	IEEE ACCESS
volume	7
page range	153238-153248
year	2019
URL	http://hdl.handle.net/10258/00010479

doi: info:doi/10.1109/ACCESS.2019.2946378

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 6, 2019, accepted October 5, 2019, date of publication October 9, 2019, date of current version October 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2946378

Team Tactics Estimation in Soccer Videos Based on a Deep Extreme Learning Machine and Characteristics of the Tactics

GENKI SUZUKI[1], (Student Member, IEEE), SHO TAKAHASHI[2], (Member, IEEE), TAKAHIRO OGAWA[1], (Senior Member, IEEE), AND MIKI HASEYAMA[1], (Senior Member, IEEE)

[1]Faculty of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan
[2]Faculty of Engineering, Hokkaido University, Sapporo 060-8628, Japan

Corresponding author: Genki Suzuki (suzuki@lmd.ist.hokudai.ac.jp)

This work was supported in part by the MIC/SCOPE under Grant #181601001, and in part by the JSPS KAKENHI under Grant JP17H01744 and Grant JP17K00148.

ABSTRACT A novel method for estimating team tactics in soccer videos based on a Deep Extreme Learning Machine (DELM) and unique characteristics of tactics is presented in this paper. The proposed method estimates the tactics of each team from players' formations and enables successful training from a limited amount of training data. Specifically, the estimation of tactics consists of two stages. First, by utilizing two DELMs corresponding to the two teams, the proposed method estimates the provisional tactics of each team. Second, the proposed method updates the team tactics based on unique characteristics of soccer tactics, the relationship between tactics of the two teams and information on ball possession. Consequently, since the proposed method estimates the team tactics that satisfy these characteristics, accurate estimation results can be obtained. In an experiment, the proposed method is applied to actual soccer videos to verify its effectiveness.

INDEX TERMS Sports video analysis, tactics estimation, deep learning, semantic analysis.

I. INTRODUCTION

In recent years, through digital broadcasting and Hybridcast [1], users have been able to find and view various multimedia contents. Furthermore, with the increase in the number of sports videos and delivery of the contents related to these videos, it is desirable to provide richer information to users. Due to the popularization of sports video distribution services such as DAZN[1] and improvement in the performance of video content analysis, many sports videos are being viewed and utilized by various audiences and by training staff of professional sports teams [2]–[4].

Many researchers have proposed methods for sports video content analysis [5]–[21]. For soccer video analysis, some fundamental methods including player tracking [5]–[7], event detection [8]–[10] and video summarization [11]–[15] have been proposed. Various new methods for advanced semantic understanding have also been proposed [16]–[21]. These methods can provide new semantic information such as movements of players [16]–[18], effective pass courses [19] and team tactics [20], [21] and they enable viewers to understand soccer videos.

Since player tracking solutions have recently been realized [22], many soccer clubs and leagues have been using them for game analysis. For example, TRACAB image tracking system[2] is used in the FIFA World Cup and used by professional leagues in Spain, Japan, and other countries [23]. It has become feasible to easily use the tracking data of players and the ball for realizing the above applications.

Soccer tactics are divided into the following three main classes: team tactics, group tactics and individual tactics [24]. Team tactics are defined as actions of the whole team and are thus very important in soccer games. Furthermore, team tactics are categorized into five groups as shown in Table 1.

The associate editor coordinating the review of this manuscript and approving it for publication was Mehedi Masud.

[1]https://www.dazn.com/

[2]TRACAB: http://chyronhego.com/sports-data/tracab/

TABLE 1. Team tactics and their brief overview [20].

Team Tactics	Attack / Defense	Overview
Retreat	Defense	Withdrawing from the opposing territory
Fore check	Defense	Pressuring the ball
Set piece	Both	Corner kick or free kick
Possession	Attack	Retaining control of the ball over longer periods of time
Swift attack	Attack	Kicking along through ball

Since the team tactics change depending on situations in soccer games, their estimation from soccer videos can provide important information about the game situation. Therefore, accurate estimation of soccer tactics is important for realizing semantic analysis of soccer videos. Therefore, it is expected that team tactics in soccer games can be estimated by utilizing the players' formation, which is determined from the players' positions and their movements.

Several methods for estimating team tactics have been proposed [20], [21]. These methods try to perform the estimation based on traditional machine learning [25] and clustering approaches [26]. With the development of deep learning technologies, it is expected that the performance of tactics estimation can be improved by introducing these technologies. However, there are the following two remaining problems.

Problem 1: Deep learning-based methods require a large number of training samples. However, there are many conditions in soccer games such as target teams, target stadiums, etc. It is difficult to prepare sufficient training samples for each condition.

Problem 2: Generally, the team tactics of the two teams in a soccer game have a strong relationship as shown in Table 2. However, since existing methods estimate the team tactics of each team independently, the estimation results for the two teams do not always satisfy this relationship. In addition, ball possession is also strongly related to team tactics, and information on ball possession should be introduced into the tactics estimation. Therefore, it is necessary to construct a new framework for estimating team tactics to solve the above problems.

TABLE 2. Relationship between team tactics of the two teams. Ret, FC, set, Poss and Sw represent retreat, fore check, set piece, possession and swift attack, respectively.

		Team A				
	Team tactics	Ret	FC	Set	Poss	Sw
Team B	Ret	-	-	-	✓	✓
	FC	-	-	-	✓	✓
	Set	-	-	✓	-	-
	Poss	✓	✓	-	-	-
	Sw	✓	✓	-	-	-

A new method for team tactics estimation based on a Deep Extreme Learning Machine (DELM) [27] and unique characteristics of team tactics in soccer games is presented in this paper. The proposed method consists of the following two stages.

1) Provisional tactics estimation based on DELM

2) Update of team tactics based on the characteristics of team tactics in soccer games

In the first stage, the proposed method estimates the provisional tactics of the two teams based on two DELMs, which are deep learning methods [28] which enables training of classifiers from limited numbers of training samples, and this provides a solution to Problem 1. We construct DELM-based classifiers using the formation features obtained from players' positions, which are provided by the TRACAB image tracking system, and their movements. In the second stage, the proposed method updates the estimation results of team tactics. Specifically, the combination of team tactics is decided on the basis of the relationship between the tactics of the two teams shown in Table 2. Finally, our method estimates the final team tactics from changes in the game situation, which is based on ball possession. The second stage provides a solution to Problem 2. Consequently, our new framework contributes to improvement in tactics estimation. Consequently, since the proposed method solves the remaining problems of the existing method, successful team tactics estimation can be expected.

It should be noted that "Set piece" shown in Table 1, one of the team tactics, which can be also regarded as an event follows the whistle after stopping the ball. It has been reported that methods for whistle sound detection [29] and ball tracking [30] realize highly accurate detection of scenes of "Set piece". Therefore, since "Set piece" can be separately detected by the existing methods, we focus on the estimation of four team tactics, "Retreat", "Fore check", "Possession" and "Swift attack", in this paper.

This paper is organized as follows. Preprocessing of our method and extraction of formation features are presented in Section II. The proposed method is presented in Section III. An overview of the preprocessing and the proposed method is shown in Fig. 1. Experimental results for verifying the effectiveness of the proposed method are shown in Section IV. Finally, concluding remarks are presented in Section V.

II. PREPROCESSING: EXTRACTION OF FORMATION FEATURES FROM SOCCER VIDEOS

In this section, we explain the extraction of formation features, for which an overview is shown in Table 3. Generally, the team tactics used in soccer games are characterized by the distribution of players on the soccer field and behaviors of individual players. The formation features include "group behavior features" and "individual behavior features", corresponding to the distribution of players on the soccer field and behaviors of individual players, respectively.

A. GROUP BEHAVIOR FEATURES

To obtain group behavior features, we focus on distances between players and their distribution except for the goalkeeper on the soccer field. Since these features depend on the team tactics, it is possible to characterize the team tactics used in soccer games. For example, each of the tactics has

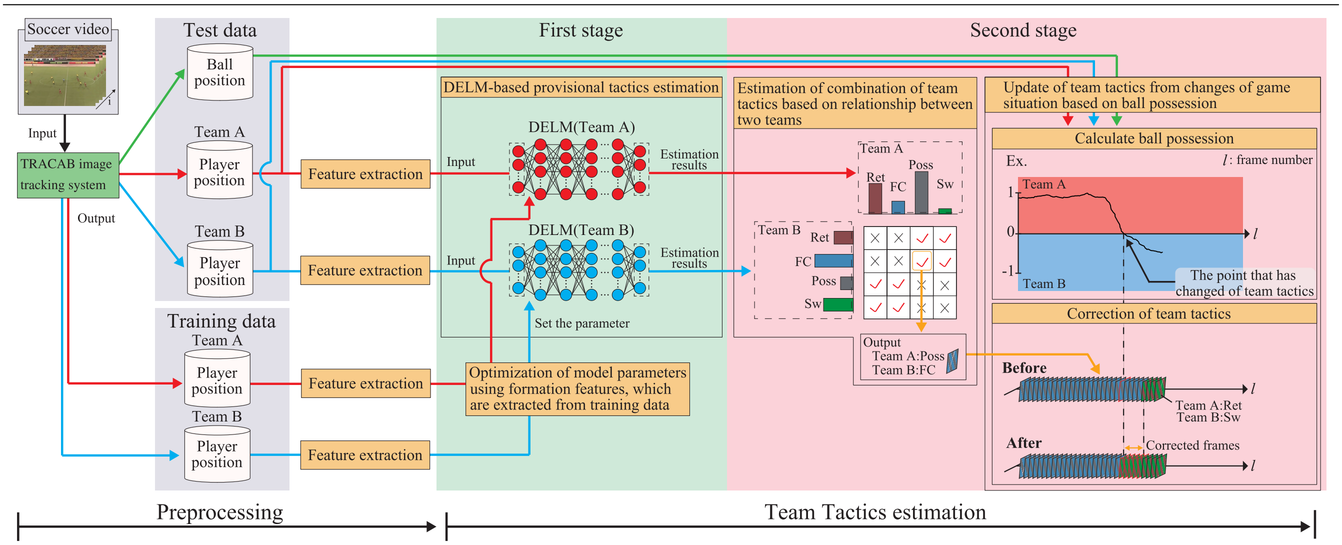


FIGURE 1. Overview of the preprocessing and the proposed method. As preprocessing, feature extraction is performed (Section II). The proposed method (Section III) consists of two stages. The first stage is provisional tactics estimation based on the DELM (III-A). The second stage is update of team tactics based on the relationship between tactics of the two teams and ball possession (III-B).

TABLE 3. Overview of formation features. Dim. represents the dimension of each feature.

Group behavior features	Dim.
Distances between players, their mean and variance	47
The mean of variation of distances between players	1
The number of players in each area on the soccer field	15
Individual behavior features	**Dim.**
All directions of players' movements, their mean and variance	24
All players' velocities, their mean and variance	12
All players' accelerations, their mean and variance	12
Total	111

a different distribution of player positions in the field, and the positions of offensive and defensive players are close. Thus, the distances between players and their distribution are effective for estimating the team tactics.

In our method, we calculate the following three kinds of group behavior features: "distances between players, their mean and variance $\boldsymbol{u}_{\text{dist}}$", "the mean of variation of distances between players u_{vdist}", and "the number of players in each area on the soccer field $\boldsymbol{u}_{\text{area}}$". These features are defined as follows:

$$\boldsymbol{u}_{\text{dist}} = [\text{dist}(1), \text{dist}(2), \ldots, \text{dist}(45), \text{mean}(\text{dist}), \text{var}(\text{dist})]^{\mathrm{T}}, \tag{1}$$

$$u_{\text{vdist}} = \text{mv}, \tag{2}$$

$$\boldsymbol{u}_{\text{area}} = [\text{area}(1), \text{area}(2), \ldots, \text{area}(15)]^{\mathrm{T}}. \tag{3}$$

In Eq. (1), dist(·) is the distance between two players except for the goalkeeper, with the total number of player combinations being 45 (${}_{10}C_2 = 45$), and mean(·) and var(·) indicate their mean and variance, respectively. Furthermore, mv is the mean of variation of distances between players, and area(m) ($m = 1, 2, \ldots, 15$) is the number of players in the divided soccer field. In this paper, the soccer field is equally divided

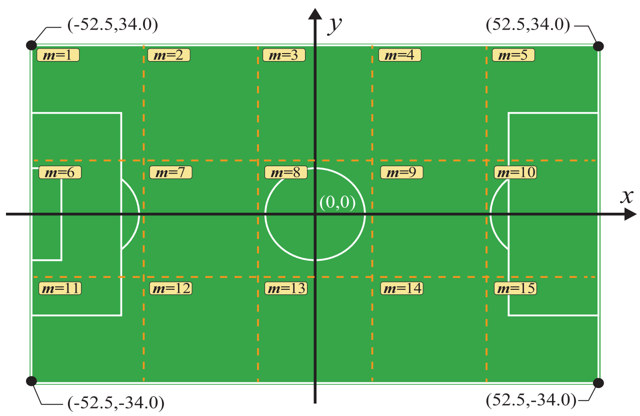


FIGURE 2. Playing field and its 15 divided regions, where x and y are axes of horizontal and vertical positions in the playing field. The size of this field follows FIFA's regulations [31].

into 15 areas as shown in Fig. 2. Finally, we can obtain $D_1(= 47 + 1 + 15)$-dimensional group behavior features.

B. INDIVIDUAL BEHAVIOR FEATURES

We use directions of players' movements, velocities and accelerations except those of the goalkeepers as individual behavior features. These features are closely related to the team tactics. For example, the directions of players' movements in defense are different from those in offence. Furthermore, in "Fore check," players often defend with rapid movements. On the other hand, in "Retreat," players defend at a steady velocity. Thus, directions of players' movements, velocities and accelerations are effective for estimating team tactics. In our method, the following three kinds of features are calculated for individual behavior features: "all directions of players' movements, their mean and variance $\boldsymbol{u}_{\text{mov}}$", "all players' velocities, their mean and variance $\boldsymbol{u}_{\text{vel}}$" and "all players' accelerations, their mean and variance $\boldsymbol{u}_{\mathrm{acc}}$''. These features are defined as follows:

$$\begin{aligned}\boldsymbol{u}_{\mathrm{mov}} = &[\mathrm{mov}_x(1), \mathrm{mov}_x(2), \ldots, \mathrm{mov}_x(10), \mathrm{mean}(\mathrm{mov}_x),\\ &\mathrm{var}(\mathrm{mov}_x), \mathrm{mov}_y(1), \mathrm{mov}_y(2), \ldots, \mathrm{mov}_y(10),\\ &\mathrm{mean}(\mathrm{mov}_y), \mathrm{var}(\mathrm{mov}_y)]^{\mathrm{T}}, \end{aligned} \tag{4}$$

$$\begin{aligned}\boldsymbol{u}_{\mathrm{vel}} = &[\mathrm{vel}(1), \mathrm{vel}(2), \ldots, \mathrm{vel}(10), \mathrm{mean}(\mathrm{vel}),\\ &\mathrm{var}(\mathrm{vel})]^{\mathrm{T}},\end{aligned} \tag{5}$$

$$\begin{aligned}\boldsymbol{u}_{\mathrm{acc}} = &[\mathrm{acc}(1), \mathrm{acc}(2), \ldots, \mathrm{acc}(10), \mathrm{mean}(\mathrm{acc}),\\ &\mathrm{var}(\mathrm{acc})]^{\mathrm{T}}.\end{aligned} \tag{6}$$

Note that

$$\mathrm{mov}_x(p) = \mathrm{pos}_x(p) - \mathrm{pos}_x^{-1}(p), \tag{7}$$

$$\mathrm{mov}_y(p) = \mathrm{pos}_y(p) - \mathrm{pos}_y^{-1}(p), \tag{8}$$

$$\mathrm{vel}(p) = \sqrt{\{\mathrm{mov}_x(p)\}^2 + \{\mathrm{mov}_y(p)\}^2}, \tag{9}$$

$$\mathrm{acc}(p) = \mathrm{vel}(p) - \mathrm{vel}^{-1}(p), \tag{10}$$

where p indicates a player ($p = 1, 2, \ldots, 10; 10$ being the total number of players except for the goalkeeper), ($\mathrm{pos}_x(\cdot)$, $\mathrm{pos}_y(\cdot)$) are horizontal and vertical positions of each player, and ($\mathrm{pos}_x^{-1}(\cdot)$, $\mathrm{pos}_y^{-1}(\cdot)$) are those in the previous time. In Eqs. (7) and (8), $\mathrm{mov}_x(\cdot)$ and $\mathrm{mov}_y(\cdot)$ are directions of players' movements, and $\mathrm{vel}(\cdot)$ and $\mathrm{acc}(\cdot)$ indicate players' velocities and accelerations, respectively, where $\mathrm{vel}^{-1}(\cdot)$ is the corresponding velocity in the previous time. Finally, we can obtain $D_2(= 24 + 12 + 12)$-dimensional individual behavior features.

III. PROPOSED METHOD FOR ESTIMATION OF TEAM TACTICS

As shown in the previous section, we obtain formation feature vectors $\boldsymbol{x}_{i,j} \in \mathbb{R}^{D_1+D_2} (i = 1, 2, \ldots, N; j \in \{\mathrm{teamA}, \mathrm{teamB}\})$ in each team, where N is the total number of frames in soccer videos. Note that $\boldsymbol{x}_{i,j}$ is a vector concatenating ''group behavior features'' and ''individual behavior features''.

In this section, we explain our method for estimation of team tactics. The right side of Fig. 1 gives an overview of the proposed method. The proposed method consists of the following two stages.

A) Provisional tactics estimation based on a DELM The proposed method estimates the team tactics of each team based on a DELM by using the formation features in soccer videos.

B) Update of estimated results of team tactics The estimation results for team tactics are updated by considering the characteristics of tactics.

The details of these two stages are shown in the following subsections.

A. FIRST STAGE: DELM-BASED ESTIMATION OF PROVISIONAL TACTICS

In this subsection, the DELM-based method for estimation of provisional team tactics for each team is presented. In recent years, methods based on a convolutional neural network (CNN) [32], [33], which is one of the deep learning techniques, have achieved more accurate recognition than that achieved by traditional machine learning-based recognition methods [25], [34]. By introducing deep learning-based methods for tactics estimation, performance improvement is expected. However, most deep learning methods including CNNs have enormous parameters to be optimized and need many training data and large computation costs. On the other hand, ELM-based methods can perform training from fewer training samples with lower computation costs [35]. Since it is difficult to prepare a large number of training samples for each condition such as each team and each stadium, the proposed method adopts a DELM for estimating the provisional team tactics for each team from the extracted formation features described in Section II.

Figure 3 shows a model structure of the DELM. The DELM consists of one input layer, K hidden layers, and one output layer, i.e., the number of layers in DELM is $K + 2$. In the training of the DELM, a weight matrix is sequentially calculated between the $(k-1)$-th layer and k-th layer. Specifically, **(I)** In the case of $k = 1, 2, \ldots, K$, the weight matrix $\boldsymbol{\alpha}_j^k (k = 1, 2, \ldots, K; j \in \mathrm{teamA}, \mathrm{teamB})$ is calculated by utilizing ELM-Auto Encoder (ELM-AE), which is an unsupervised learning method, in each hidden layer. **(II)** In the case of $k = K + 1$, the weight matrix $\boldsymbol{\alpha}_j^{K+1} (j \in \mathrm{teamA}, \mathrm{teamB})$ is calculated in the same manner as ELM, which is a supervised learning method [28]. In the rest of this section, training of the DELM, i.e., calculation of the weight matrix is explained.

(I) In the case of $k = 1, 2, \ldots, K$

The relationship between the k-th hidden layer's output $\boldsymbol{H}_j^k = [\boldsymbol{h}_{1,j}^k, \boldsymbol{h}_{2,j}^k, \ldots, \boldsymbol{h}_{N,j}^k]^{\mathrm{T}} \in \mathbb{R}^{N \times L^k}$ and the $(k-1)$-th hidden layer's output $\boldsymbol{H}_j^{k-1} \in \mathbb{R}^{N \times L^{k-1}}$ can be obtained as follows:

$$\boldsymbol{H}_j^k = g\left(\boldsymbol{H}_j^{k-1}\left(\boldsymbol{\alpha}_j^k\right)^{\mathrm{T}}\right), \tag{11}$$

where $\boldsymbol{\alpha}_j^k \in \mathbb{R}^{L^k \times L^{k-1}}$ is the weight matrix between the (k-1)-th and k-th hidden layers, and $g(\cdot)$ is the activation function. Note that L^k is the number of nodes in the k-th hidden layer. The input layer is expressed by $k - 1 = 0$, and $\boldsymbol{H}_j^0$ consists of the obtained feature vectors $\boldsymbol{x}_{i,j}$ described in the previous subsection. In order to calculate the weight matrix $\boldsymbol{\alpha}_j^k$, the proposed method generates ELM-AE in each layer. Given an input vector $\boldsymbol{h}_{i,j}^{k-1}$, the outputs $\boldsymbol{w}_{i,j}^k$ of the hidden layer in ELM-AE can be obtained as

$$\boldsymbol{w}_{i,j}^k = g\left(\boldsymbol{A}^k \boldsymbol{h}_{i,j}^{k-1} + \boldsymbol{b}^k\right), \tag{12}$$

where $\boldsymbol{A}^k = [\boldsymbol{a}_1^k, \boldsymbol{a}_2^k, \ldots, \boldsymbol{a}_{L^k}^k]^{\mathrm{T}} \in \mathbb{R}^{L^k \times L^{k-1}}$ is an orthogonal random weight matrix in ELM-AE, and $\boldsymbol{b}^k = [b_1^k, b_2^k, \ldots, b_{L^k}^k]^{\mathrm{T}} \in \mathbb{R}^{L^k}$ is a random bias vector in ELM-AE. By using ELM-AE's input matrix $\boldsymbol{H}_j^{k-1}$ and its hidden layer output matrix $\boldsymbol{W}_j^k = [\boldsymbol{w}_{j,1}^k, \boldsymbol{w}_{j,2}^k, \ldots, \boldsymbol{w}_{j,N}^k]^{\mathrm{T}} \in \mathbb{R}^{N \times L^k}$,

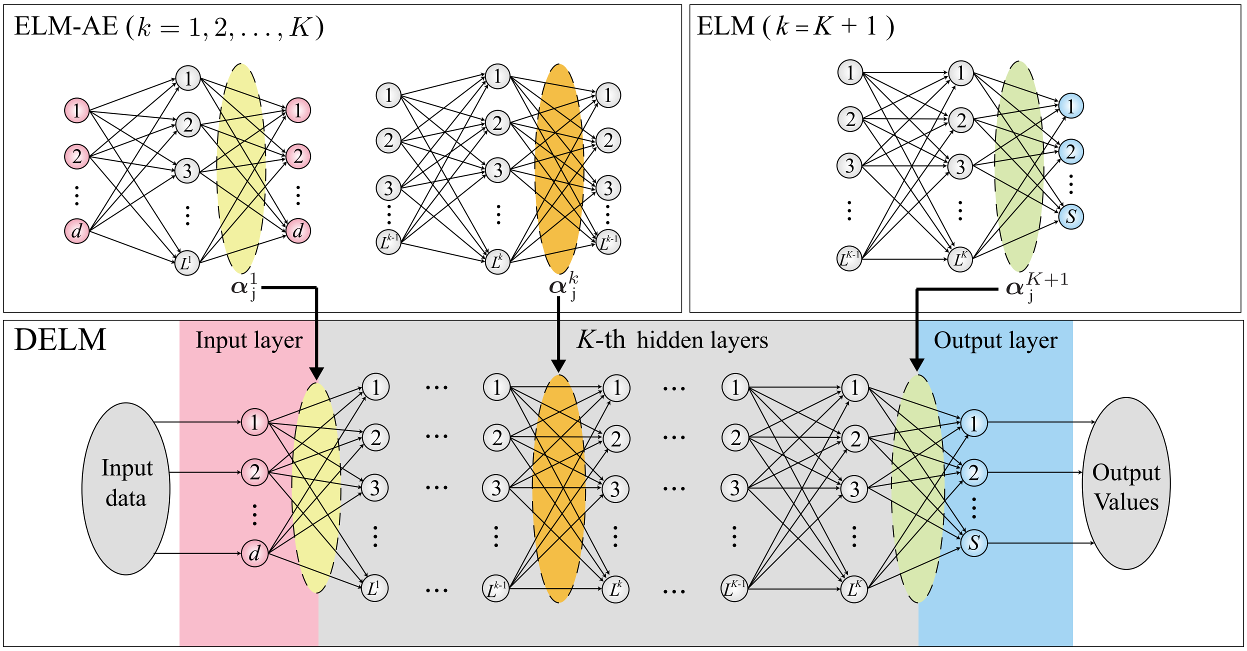


FIGURE 3. Model structure of the DELM. ELM-AE weights α_j^1 and α_j^k are the first and k-th layers' weights of the DELM, respectively. ELM weights α_j^{K+1} are the $(K+1)$-th layer's weights of the DELM. The weights of each DELM layer are calculated by using ELM-AE ($k = 1, 2, \ldots, K$) or ELM($k = K + 1$), and estimation results v_j are obtained by inputting x_j.

the output weight matrix $\boldsymbol{\alpha}_j^k$ is calculated by the following two patterns.

(**I-A**) In the case of $L^k \neq L^{k-1}$
The output weight matrix $\boldsymbol{\alpha}_j^k$ is obtained as follows:

$$\boldsymbol{\alpha}_j^k = \left(\frac{\boldsymbol{I}}{C_1} \sum_{l^k=1}^{L^k} \mathrm{KL}(\rho \| \hat{\rho}_{l^k}) + \left(\boldsymbol{W}_j^k\right)^{\mathrm{T}} \boldsymbol{W}_j^k \right)^{-1} \times \left(\boldsymbol{W}_j^k\right)^{\mathrm{T}} \boldsymbol{H}_j^{k-1}, \tag{13}$$

$$\mathrm{KL}(\rho \| \hat{\rho}_{l^k}) = \rho \log \frac{\rho}{\hat{\rho}_{l^k}} + (1-\rho) \log \frac{1-\rho}{1-\hat{\rho}_{l^k}}, \tag{14}$$

where $\mathrm{KL}(\rho \| \hat{\rho}_{l^k})$ is the KL divergence, ρ is a parameter of activation ($\rho = 0.05$), $\hat{\rho}_{l^k}$ is the average activation of each hidden layer's node l^k of ELM-AE, $\boldsymbol{I}$ is the identity matrix, and C_1 is a regularization parameter.

(**I-B**) In the case of $L^k = L^{k-1}$
In this case, the output weight matrix $\boldsymbol{\alpha}_j^k$ is obtained by solving the following orthogonal procrustes problem:

$$\boldsymbol{\alpha}_j^k = \underset{\boldsymbol{\Omega}}{\operatorname{argmin}} ||\boldsymbol{\Omega}(\boldsymbol{H}_j^{k-1})^{\mathrm{T}} - (\boldsymbol{W}_j^k)^{\mathrm{T}}||_F \quad \text{subject to} \quad \boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{\Omega} = \boldsymbol{I}, \tag{15}$$

where $\|\cdot\|_F$ is the Frobenius norm. Specifically, by solving Eq. (15) based on the following equation via singular value decomposition based on [36], $\boldsymbol{\alpha}_j^k$ can be obtained.

$$\boldsymbol{\alpha}_j^k = \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}, \tag{16}$$

$$(\boldsymbol{W}_j^k)^{\mathrm{T}}\boldsymbol{H}_j^{k-1} = \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{V}^{\mathrm{T}}, \tag{17}$$

where $\boldsymbol{U} \in \mathbb{R}^{L^k \times L^k}$ and $\boldsymbol{V} \in \mathbb{R}^{L^{k-1} \times L^{k-1}}$ are orthogonal matrices, $\boldsymbol{\Gamma} \in \mathbb{R}^{L^k \times L^{k-1}}$ is a singular value matrix, and $(\boldsymbol{\alpha}_j^k)^{\mathrm{T}}\boldsymbol{\alpha}_j^k = \boldsymbol{I}$. Consequently, in the proposed method, each layer's output weight matrix $\boldsymbol{\alpha}_j^k$ is obtained.

(**II**) In the case of $k = K + 1$ The output weight matrix $\boldsymbol{\alpha}_j^k$ between the K-th hidden layer and the output layer is calculated by ELM-based supervised learning. Specifically, $\boldsymbol{\alpha}_j^k$ can be obtained as

$$\boldsymbol{\alpha}_j^k = \boldsymbol{T}_j \boldsymbol{H}_j^k \left(\frac{\boldsymbol{I}}{C_2} + \boldsymbol{H}_j^k (\boldsymbol{H}_j^k)^{\mathrm{T}} \right)^{-1}, \tag{18}$$

where $\boldsymbol{T}_j = [\boldsymbol{t}_{j,1}, \boldsymbol{t}_{j,2}, \ldots, \boldsymbol{t}_{j,N}] \in \mathbb{R}^{S \times N}$, $\boldsymbol{t}_{j,n} = [t_{i,j,1}, t_{i,j,2}, \ldots, t_{i,j,S}]^{\mathrm{T}}$ includes training labels, and C_2 is a regularization parameter. If a true class label of $\boldsymbol{h}_{i,j}^k$ is s ($s = 1, 2, \ldots, S$), $t_{i,j,s} = 1$, where $S(= 4)$ is the number of team tactics. On the other hand, the other elements are zeros.

In this way, we can perform training of the DELM-based classifiers. In the test phase, by inputting new formation feature vectors $\boldsymbol{x}_{i,j}$ to the above DELM, output values $v_{i,j}^{\mathrm{Ret}}$, $v_{i,j}^{\mathrm{FC}}$, $v_{i,j}^{\mathrm{Poss}}$, $v_{i,j}^{\mathrm{Sw}}$ corresponding to each of the tactics from the output layer of the DELM are obtained.

B. SECOND STAGE: UPDATE OF TEAM TACTICS CONSIDERING THE CHARACTERISTICS OF TACTICS

In this subsection, we explain the update of team tactics considering the characteristics of tactics in soccer games. The proposed method tries to update the team tactics more accurately by applying two procedures to the results obtained in III-A. Specifically, these procedures are 1) estimation of the combination of team tactics based on the relationship between tactics of the two teams and 2) update of team tactics from changes in the game situation based on ball possession. In the first procedure, since the relationship between team tactics of the two teams is limited to the eight patterns, as shown in Table 2, it is possible to prevent the occurrence of a failure case, e.g., a case in which the tactics of both teams become defensive. In the second procedure, it is possible to accurately determine temporal changes in team tactics since temporal changes in the game situation can be found from ball possession data. In the rest of this subsection, we show the details of these two procedures.

1) Estimation of combination of team tactics based on the relationship between tactics of the two teams

We show how to estimate the combination of team tactics based on the relationship between tactics of the two teams. In the proposed method, we obtain eight kinds of values $z_c(c \in \text{Comb})$ corresponding to the possible combinations except for "Set piece" in the second stage of Fig. 1, where Comb = {(Ret, Poss), (Ret, Sw), (FC, Poss), (FC, Sw), (Poss, Ret), (Poss, FC), (Sw, Ret), (Sw, FC)}. The values of z_c are obtained by adding output values of the corresponding team tactics of the two teams in DELMs. For example, to obtain the value of the combination (Ret, Sw), i.e., "Retreat" in team A and "Swift attack" in team B, we add $v^{\text{Ret}}_{i,\text{teamA}}$ to $v^{\text{Sw}}_{i,\text{teamB}}$ for obtaining $z_{(\text{Ret},\text{Sw})}$. Finally, the proposed method obtains the optimal combination of team tactics by utilizing the following equation:

$$\text{Label} = \arg\max_{c \in \text{Comb}} z_c. \tag{19}$$

In the proposed method, voting is performed for the results of the combination of team tactics to correct its instantaneous error. Specifically, voting is performed for the estimation results in Eq. (17) for M frames before and after the target frame. Consequently, the proposed method obtains the estimation results based on the relationship between tactics of the two teams according to Table 2.

2) Update of team tactics from changes in the game situation based on ball possession

We estimate points in the team tactics that have changed based on ball possession and correct the estimation results by using these points that have changed. The method for correcting estimation results based on ball possession is shown in Fig. 4. First, we calculate the ball possession Ball_l of target frame l by using the ball and players' position data, where Ball_l represents the degree of ball possession for one team. Next, we regard frame l of the point that has changed when the ball retention rate exceeds or falls below zero. Finally, the proposed method obtains the estimation results by correcting team tactics of the neighboring frames based on the point in the team tactics that has changed. The calculation of Ball_l is described in detail below.

Ball_l is calculated on the basis of ball possession in each target frame. A player that has the ball satisfies the following four conditions.

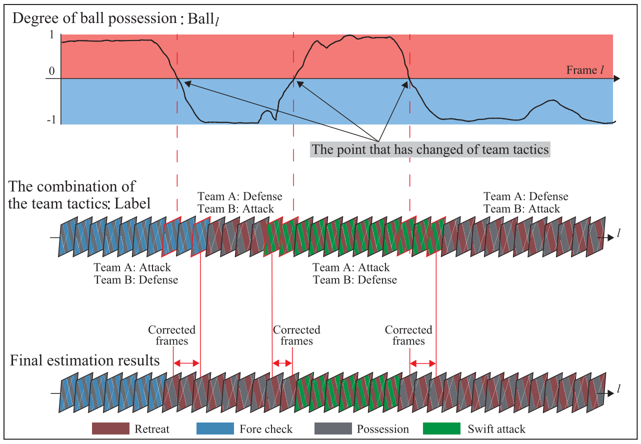


FIGURE 4. Method for correcting the estimation results based on ball possession.

Condition 1:
The player with the ball is the closest to the ball among all players.

Condition 2:
The distance between the nearest player and the ball is within a fixed distance r[m].

Condition 3:
$\left||\boldsymbol{b}_l| - |\boldsymbol{b}_{l-1}|\right| < \text{Th}_{\text{min}}$ or $\left||\boldsymbol{b}_l| - |\boldsymbol{b}_{l-1}|\right| > \text{Th}_{\text{max}}$.

Condition 4:
$\frac{\boldsymbol{b}_l \cdot \boldsymbol{b}_{l-1}}{|\boldsymbol{b}_l||\boldsymbol{b}_{l-1}|} < \text{Th}_{\text{a}}$.

Note that $\boldsymbol{b}_l$ is a vector representing the difference in ball position between the l-th frame and $(l+1)$-th frame. Th_{min}, Th_{max} and Th_{a} are thresholds.

Ball_l is calculated by using the estimation results for ball retention as follows:

$$\text{Ball}_l = \frac{1}{T} \sum_{l'=l-T}^{l} f(l'), \tag{20}$$

$$f(l') = \begin{cases} 1 & \text{if } B_{l'} = \text{teamA} \\ -1 & \text{Otherwise} \end{cases}, \tag{21}$$

where $B_{l'}$ is the ball retention in the frame l'.

Thus, the estimation of tactics based on the DELM considering the unique characteristics of team tactics in soccer games become feasible by the proposed method.

IV. EXPERIMENTAL RESULTS

In this section, we verify the effectiveness of the proposed method using actual soccer videos. In IV-A, an outline of the experiment is explained. Experimental results are shown in IV-B.

A. EXPERIMENTAL CONDITIONS

Soccer videos and tracking data for players and balls (4,611 seconds, 5 fps) were used in the experiment. Table 4 shows details of the dataset. In the dataset, a target team performed soccer games using Retreat, Fore check, Possession and Swift attack. We conducted the experiment using many more frames of videos than those used in other methods

TABLE 4. Amount of inspection data used in the experiment.

Team tactics	Num. of frames
Retreat	1,541
Fore check	613
Possession	2,045
Swift attack	412
Total	4,611

such as [9], [18]. The ground truth of the team tactics was determined by subjects who had ten years of experience as soccer players.

We used a five-fold cross-validation for model evaluation. Then we used Recall, Precision and F-measure for evaluation criteria defined as

$$\text{Recall} = \frac{\text{Num. of correctly estimated team tactics}}{\text{Num. of true team tactics}}, \tag{22}$$

$$\text{Precision} = \frac{\text{Num. of correctly estimated team tactics}}{\text{Num. of estimated team tactics}}, \tag{23}$$

$$\text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}. \tag{24}$$

In order to confirm the effectiveness of the proposed method, the following methods were adopted for comparative methods (Comp. 1 - Comp. 10).

- Comp. 1 is a baseline method and our recently reported method [20] based on Multiple Kernel Fuzzy C-Means (MKFC) [26] using formation features.
- Comp. 2 is a baseline method and our recently reported method [21] based on Decision Level Fusion [37]. Specifically, Comp. 2 estimates team tactics based on a Support Vector Machine (SVM) [25] using both formation features and audio-visual features.
- Comp. 3 is an estimation method in which the MKFC of Comp. 1 is replaced by Random Forest [38].
- Comp. 4 is an estimation method in which the MKFC of Comp. 1 is replaced by K-Nearest Neighbor algorithm (K-NN) [39]
- Comp. 5 applies the same framework as that of the proposed approach to different features. Specifically, the formation features of the proposed method are replaced by the players' positions.
- Comp. 6 is an estimation method in which the DELM of the proposed method is replaced by SVM.
- Comp. 7 is an estimation method in which the DELM of the proposed method is replaced by ELM [28] like Comp. 5.
- Comp. 8 is an estimation method including the first stage of the proposed method without applying the second stage of the proposed method.
- Comp. 9 estimates team tactics based on a fine-tuned VGG16 model [40], which is one of the benchmarking CNN-based approaches.
- Comp. 10 is a method for estimation of team tactics based on the fine-tuned InceptionV3 model [41], which is also one of the CNN-based approaches.

TABLE 5. Details of parameters used in the proposed method.

Details	Parameter	Value
Num. of input nodes	d	111
Num. of hidden layers	K	6
Regularization parameter	C_1	-2^{10}
	C_2	-2^{10}
Num. of hidden nodes(Team A)	L^1	131
	L^2	131
	L^3	171
	L^4	131
	L^5	111
	L^6	151
Num. of hidden nodes(Team B)	L^1	131
	L^2	111
	L^3	171
	L^4	171
	L^5	111
	L^6	151
Num. of output nodes	S	4
Frame size of the voting process	M	21
Parameters of ball possession	r	2.0
	$\mathrm{Th}_{\mathrm{min}}$	0.3
	$\mathrm{Th}_{\mathrm{max}}$	1.2
	Th_{a}	$\cos\frac{\pi}{6}$
	T	15

The validity of these comparative methods is described as follows.

Proposed method vs Comp. 1 and Comp. 2

We verify that the proposed method is more effective than the baseline method.

Proposed method vs Comp. 3 and Comp. 4

We verify that the proposed method is more effective than traditional machine learning-based methods.

Proposed method vs Comp. 5

We verify the effectiveness of the use of formation features.

Proposed method vs Comp. 6 and Comp. 7

We verify the effectiveness of the use of a DELM in the first stage.

Proposed method vs Comp. 8

We verify the effectiveness of the use of the second stage.

Proposed method vs Comp. 9 and Comp. 10

We verify that the proposed method is more effective than general CNN-based methods.

Details of the parameters used in the proposed method are shown in Table 5. The parameters in each method were determined in such a way that the performance of each method was the highest. In Comp. 9 and Comp. 10, to make fair comparison, we used the Tensorflow [42] implementation and download the VGG16 model and InceptionV3 model. The training was conducted at a maximum of 50 epochs until the error was at least 0.2. Finally, the best network was selected based on performance measured through the accuracy of the validation dataset.

TABLE 6. Comparison of team tactics estimation by our method and by the comparative methods.

Team tactics	Ours R	Ours P	Ours F	Comp. 1 R	Comp. 1 P	Comp. 1 F	Comp. 2 R	Comp. 2 P	Comp. 2 F
Retreat	**0.928**	**0.917**	**0.923**	0.780	0.394	0.523	0.683	0.670	0.676
Fore check	0.871	**0.830**	**0.850**	0.408	0.658	0.504	0.611	0.651	0.630
Possession	**0.928**	**0.945**	**0.936**	0.653	0.471	0.547	0.699	0.720	0.709
Swift attack	0.791	**0.812**	**0.801**	0.112	0.565	0.187	0.637	0.644	0.640
Average	**0.880**	**0.876**	**0.878**	0.488	0.522	0.440	0.658	0.671	0.664

Team tactics	Comp. 3 R	Comp. 3 P	Comp. 3 F	Comp. 4 R	Comp. 4 P	Comp. 4 F	Comp. 5 R	Comp. 5 P	Comp. 5 F
Retreat	0.657	0.672	0.665	0.600	0.616	0.608	0.877	0.825	0.850
Fore check	0.660	0.628	0.644	0.605	0.581	0.593	0.633	0.779	0.699
Possession	0.671	0.672	0.671	0.617	0.612	0.615	0.881	0.827	0.854
Swift attack	0.664	0.647	0.656	0.610	0.592	0.601	0.643	0.752	0.693
Average	0.663	0.655	0.669	0.608	0.600	0.604	0.759	0.796	0.774

Team tactics	Comp. 6 R	Comp. 6 P	Comp. 6 F	Comp. 7 R	Comp. 7 P	Comp. 7 F	Comp. 8 R	Comp. 8 P	Comp. 8 F
Retreat	0.878	0.860	0.869	0.856	0.837	0.847	0.888	0.853	0.870
Fore check	0.776	0.766	0.771	0.805	0.755	0.779	0.807	0.801	0.804
Possession	0.822	0.854	0.838	0.868	0.894	0.881	0.878	0.899	0.888
Swift attack	0.789	0.804	0.796	0.676	0.843	0.751	0.747	0.786	0.766
Average	0.816	0.821	0.818	0.802	0.832	0.814	0.830	0.835	0.832

Team tactics	Comp. 9 R	Comp. 9 P	Comp. 9 F	Comp. 10 R	Comp. 10 P	Comp. 10 F
Retreat	0.790	0.835	0.812	0.697	0.873	0.775
Fore check	0.771	0.683	0.724	**0.880**	0.427	0.575
Possession	0.819	0.831	0.825	0.695	0.876	0.775
Swift attack	**0.807**	0.638	0.713	0.692	0.860	0.767
Average	0.797	0.747	0.768	0.741	0.759	0.723

B. PERFORMANCE EVALUATION

Recall, Precision and F-measure of all methods are shown in Table 6. From this table, it is confirmed that the proposed method outperforms the comparative methods. Specifically, our method outperforms the baseline and traditional machine learning-based methods (Comp. 1 - Comp. 4). Next, by comparing the results of the proposed method with those of Comp. 5, the effectiveness of formation features is confirmed. Furthermore, by comparing the results of the proposed method with those of Comp. 6 and Comp. 7, the effectiveness of the use of a DELM in the first stage can be confirmed. Moreover, since the performance of the proposed method is higher than that of Comp. 8, it is confirmed that the introduction of the second stage is effective. Finally, by comparing the results of the proposed method with those of Comp. 9 and Comp. 10, it is confirmed that our method is superior to other deep learning methods including very deep networks. The results indicate that a CNN is not suitable for a small amount of training samples. Therefore, the proposed method solves Problem 1. In addition, it has been reported in [27] that a DELM provides better performance than that of an SVM, and the same tendency was confirmed in the experiment.

Some of the estimation results of the proposed method and the comparative methods are shown in Fig. 5. From the results obtained, we can confirm that the proposed method achieves the most accurate estimation, the results being most similar to the ground truth. Specifically, Comp. 8 estimates the same tactics in the two teams, whereas the proposed method estimates the tactics in the two teams without contradiction. From the above, the estimation performance can be improved by using the relationship between team tactics of the two teams. Furthermore, confusion matrices for estimation results of the proposed method and Comp. 8 are shown in Tables 7 and 8. From these tables, we can see that introduction of the second stage can improve the estimation performance. Specifically, it becomes possible to estimate "Retreat" and "Possession" correctly. Generally, these team tactics are similar formations. The proposed method correctly updates the estimation results of these team tactics by utilizing ball possession. In this way, the proposed method solves Problem 2. Therefore,

TABLE 7. Confusion matrix obtained from the estimation results before performing calculation of the optimal combination of team tactics (second stage) in the proposed method. These results correspond to those of Comp. 8. Gt represents the ground truth.

		Estimation results			
		Ret	FC	Poss	Sw
Gt	Ret	1,396	48	110	19
	FC	31	510	39	52
	Poss	152	67	1,720	20
	Sw	57	12	44	334

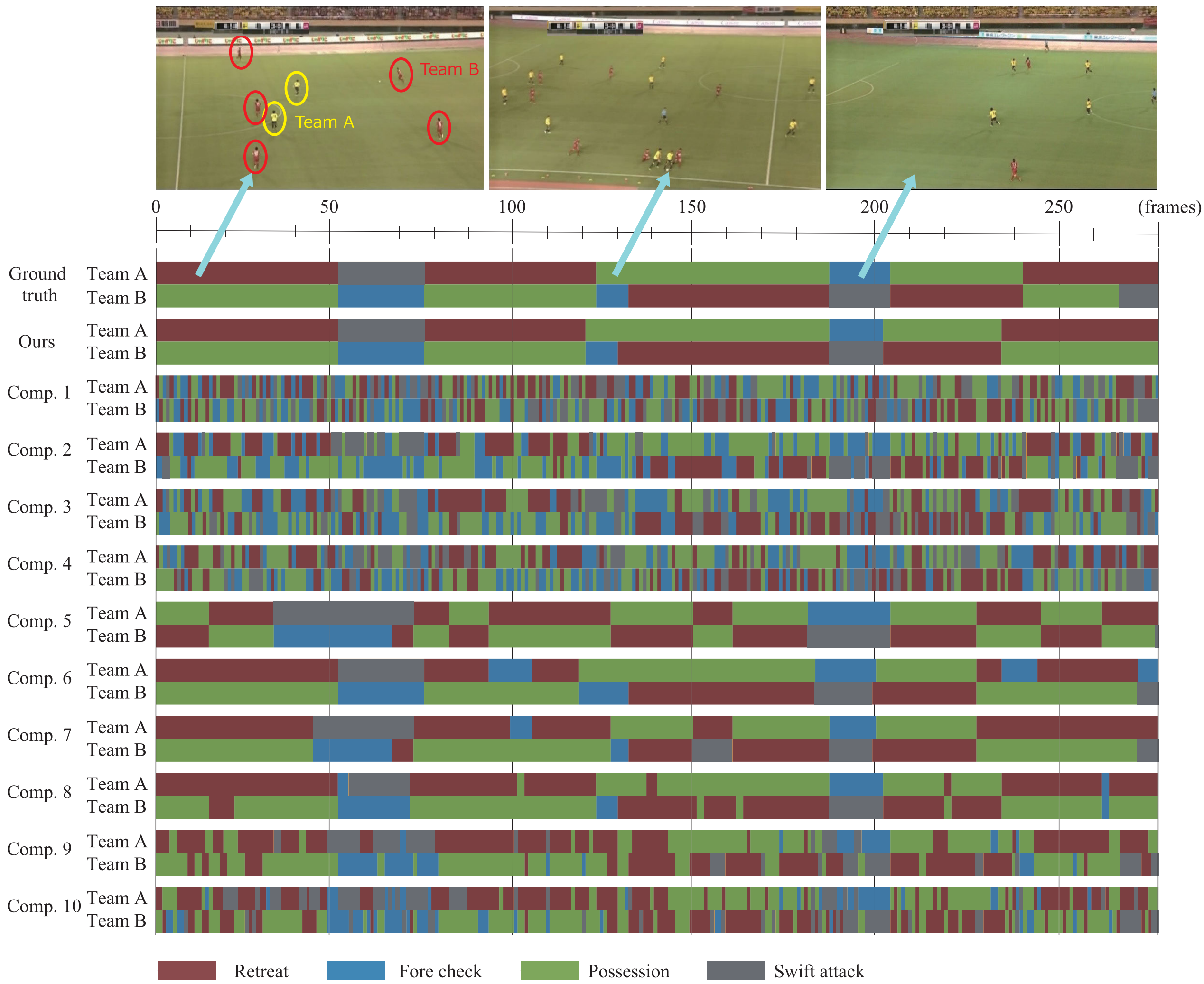


FIGURE 5. Results of tactics estimation obtained by the proposed method and the comparative methods.

TABLE 8. Confusion matrix obtained from the estimation results after performing calculation of the optimal combination of team tactics (second stage) in the proposed method. These results correspond to those of our method. Gt represents the ground truth.

		Estimation results			
		Ret	FC	Poss	Sw
Gt	Ret	1,425	43	50	17
	FC	9	527	22	47
	Poss	80	54	1,897	14
	Sw	40	11	38	337

the novelties of the proposed method strongly contribute to the successful estimation of team tactics.

V. CONCLUSION

A method for estimating team tactics in soccer videos based on a DELM and unique characteristics of tactics is presented in this paper. The proposed method estimates provisional tactics of the two teams based on two DELMs. The final team tactics are estimated on the basis of the relationship between tactics of the two teams and ball possession. The proposed method realizes accurate estimation of team tactics. Experimental results show that the proposed method outperforms comparative methods, and the novel approach in our method contributes to the improvement in performance.

In the future, we will extend our system to other sports applications. Since soccer has more players than other sports and its movement is complicated, our method can be applied to tactics analysis of many sports. For example, it can be applied to tactics analysis of field sports related to players' formation such as basketball, rugby, American football, and ice hockey. Moreover, the proposed method can be applied to the analysis of general group behavior since it is intended to analyze players in each group. Various methods for general group behavior have also been proposed, e.g., detection of suspicious persons in department stores and stations [43], measurement of human flow at evacuation behavior [44], and analysis of traffic jams [45], [46]. By applying our method to the analysis of group behavior, a wide range of possibilities can be expected.

REFERENCES

[1] H. Ohmata, M. Takechi, S. Mitsuya, K. Otsuki, A. Baba, K. Matsumura, K. Majima, and S. Sunasaki, "Hybridcast: A new media experience by integration of broadcasting and broadband," in *Proc. ITU Kaleidoscope Building Sustain. Communities*, Apr. 2013, pp. 1–8.

[2] M. Partington, C. J. Cushion, E. Cope, and S. Harvey, "The impact of video feedback on professional youth football coaches' reflection and practice behaviour: A longitudinal investigation of behaviour change," *Reflective Pract.*, vol. 16, no. 5, pp. 700–716, 2015.

[3] W. G. Taylor, P. Potrac, L. J. Nelson, L. Jones, and R. Groom, "An elite hockey player's experiences of video-based coaching: A poststructuralist reading," *Int. Rev. Sociol. Sport*, vol. 52, no. 1, pp. 112–125, 2017.

[4] M. Milbrath, P. Stoepker, and J. M. Krause, "Video analysis tools for the assessment of running efficiency," *Track Cross Country J.*, vol. 2, no. 4, pp. 279–283, 2016.

[5] J. Liu, X. Tong, W. Li, T. Wang, Y. Zhang, and H. Wang, "Automatic player detection, labeling and tracking in broadcast soccer video," *Pattern Recognit. Lett.*, vol. 30, no. 2, pp. 103–113, 2009.

[6] K. Soomro, S. Khokhar, and M. Shah, "Tracking when the camera looks away," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV) Workshops*, Dec. 2015, pp. 25–33.

[7] H. Kataoka, K. Hashimoto, and Y. Aoki, "Player position estimation by monocular camera for soccer video analysis," in *Proc. SICE Annu. Conf.*, Sep. 2011, pp. 1985–1990.

[8] C. Poppe, S. D. Bruyne, and R. Van de Walle, "Generic architecture for event detection in broadcast sports video," in *Proc. 3rd Int. Workshop Automated Inf. Extraction Media Prod.*, 2010, pp. 51–56.

[9] T. Tsunoda, Y. Komori, M. Matsugu, and T. Harada, "Football action recognition using hierarchical LSTM," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 99–107.

[10] G. Zhu, C. Xu, Q. Huang, Y. Rui, S. Jiang, W. Gao, and H. Yao, "Event tactic analysis based on broadcast sports video," *IEEE Trans. Multimedia*, vol. 11, no. 1, pp. 49–67, Jan. 2009.

[11] Y.-M. Su and C.-H. Hsieh, "A novel model-based segmentation approach to extract caption contents on sports videos," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2006, pp. 1829–1832.

[12] M. A. Refaey, W. Abd-Almageed, and L. S. Davis, "A logic framework for sports video summarization using text-based semantic annotation," in *Proc. 3rd Int. Workshop Semantic Media Adaptation Personalization*, Dec. 2008, pp. 69–75.

[13] G. Li, S. Ma, and Y. Han, "Summarization-based video caption via deep neural networks," in *Proc. 23rd Int. Conf. Multimedia*, 2015, pp. 1191–1194.

[14] M. Tavassolipour, M. Karimian, and S. Kasaei, "Event detection and summarization in soccer videos using Bayesian network and Copula," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 2, pp. 291–304, Feb. 2014.

[15] N. Babaguchi, Y. Kawai, and T. Kitahashi, "Event based indexing of broadcasted sports video by intermodal collaboration," *IEEE Trans. Multimedia*, vol. 4, no. 1, pp. 68–75, Mar. 2002.

[16] C. Perin, R. Vuillemot, and J.-D. Fekete, "SoccerStories: A kick-off for visual soccer analysis," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2506–2515, Dec. 2013.

[17] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Action classification in soccer videos with long short-term memory recurrent neural networks," in *Artificial Neural Networks*. 2010, pp. 154–159.

[18] J. Perl, A. Grunz, and D. Memmert, "Tactics analysis in soccer-an advanced approach," *Int. J. Comput. Sci. Sport*, vol. 12, no. 1, pp. 33–44, 2013.

[19] S. Takahashi and M. Haseyama, "Active grid-based pass region estimation from multiple frames of broadcast soccer videos," *ITE Trans. Media Technol. Appl.*, vol. 1, no. 3, pp. 220–225, 2013.

[20] S. Ohnuki, S. Takahashi, T. Ogawa, and M. Haseyama, "Soccer video segmentation based on team tactics estimation method," in *Proc. Int. Workshop Adv. Image Technol.*, 2013, pp. 692–695.

[21] G. Suzuki, S. Takahashi, T. Ogawa, and M. Haseyama, "Decision level fusion-based team tactics estimation in soccer videos," in *Proc. IEEE 5th Global Conf. Consum. Electron.*, Oct. 2016, pp. 58–59.

[22] W. Gregson and M. Littlewood, *Science in Soccer: Translating Theory into Practice*. London, U.K.: Bloomsbury Publishing, 2018.

[23] C. Kenneth and R. A. Daniel, "The application of sports technology and sports data for commercial purposes," in *The Use of Technology in Sport-Emerging Challenges*. London, U.K.: IntechOpen, 2018.

[24] J. Buschmann, H. Bussmann, and K. Pabst, *Coordination: A New Approach to Soccer Coaching*. Aachen, Germany: Meyer Meyer Sport, 2002.

[25] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[26] H. C. Huang, Y. Y. Chuang, and C. S. Chen, "Multiple Kernel fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 1, pp. 120–134, Feb. 2012.

[27] E. Cambria , G. B. Huang, L. L. C. Kasun, H. Zhou, C. M. Vong, J. Lin, J. Yin, Z. Cai, Q. Liu, K. Li, V. C. M. Leung, L. Feng, Y. S. Ong, M. H. Lim, X. Yang, K. Mao, B. S. Oh, H. Yu, Y. Chen, and J. Liu, "Extreme learning machines [trends & controversies]," *IEEE Intell. Syst.*, vol. 28, no. 6, pp. 30–59, Nov./Dec. 2013.

[28] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2, Jul. 2004, pp. 985–990.

[29] J. Wang, C. Xu, E. Chng, K. Wah, and Q. Tian, "Automatic replay generation for soccer video broadcasting," in *Proc. Int. Conf. Multimedia*, 2004, pp. 32–39.

[30] X. Yu, C. Xu, H. W. Leong, Q. Tian, Q. Tang, and K. W. Wan, "Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video," in *Proc. 11th Int. Conf. Multimedia*, 2003, pp. 11–20.

[31] *Laws of the Game*, Zurich, Switzerland, Int. Football Assoc. Board, 2018.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 2, 2012, pp. 1097–1105.

[33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[34] J. Sánche and F. Perronnin, "High-dimensional signature compression for large-scale image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1665–1672.

[35] S. Ding, N. Zhang, X. Xu, L. Guo, and J. Zhang, "Deep extreme learning machine and its application in EEG classification," *Math. Problems Eng.*, vol. 2015, Nov. 2014, Art. no. 129021.

[36] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 1330–1334, Dec. 2000.

[37] C. V. Rayker, S. Yu, L. H. Zhao, A. Jerebko, C. Florin, G. H. Valadez, L. Bogoni, and L. Moy, "Supervised learning from multiple experts: Whom to trust when everyone lies a bit," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 889–896.

[38] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[39] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[41] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[42] M. Abadi *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: https://arxiv.org/abs/1603.04467

[43] R. Arroyo, J. J. Yebes, L. M. Bergasa, I. G. Daza, and J. Almazán, "Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls," *Expert Syst. With Appl.*, vol. 42, no. 21, pp. 7991–8005, 2015.

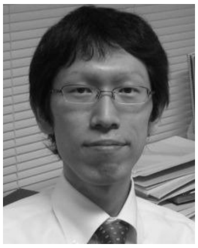
[44] T. Sano, M. Yajima, H. Kadokura, and A. Sekizawa, "Human behavior in a staircase during a total evacuation drill in a high-rise building," *Fire Mater.*, vol. 41, no. 4, pp. 375–386, 2017.

[45] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. van de Wetering, "Visual traffic jam analysis based on trajectory data," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2159–2168, Dec. 2013.

[46] A. M. de Souza, R. S. Yokoyama, G. Maia, A. Loureiro, and L. Villas, "Real-time path planning to prevent traffic jam through an intelligent transportation system," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2016, pp. 726–731.

GENKI SUZUKI (S'16) received the B.S. and M.S. degrees in electronics and information engineering from Hokkaido University, Japan, in 2017 and 2019, respectively, where he is currently pursuing the Ph.D. degree with the Graduate School of Information Science and Technology. His research interest includes sports video analysis. He is a Student Member of the IEICE.

SHO TAKAHASHI (S'09–M'13) received the B.S., M.S., and Ph.D. degrees in electronics and information engineering from Hokkaido University, Japan, in 2008, 2010, and 2013, respectively. He joined the Graduate School of Information Science and Technology, Hokkaido University, as an Assistant Professor, in 2013. He was an Associate Professor of the Education and Research Center for Mathematical and Data Science, Hokkaido University, from 2017 to 2018. He was a Visiting Researcher with the Media Integration and Communication Center, University of Florence, from 2018 to 2019. He is currently an Associate Professor with the Faculty of Engineering, Hokkaido University. His research interest includes semantic analysis and visualization in various data and its applications. He is a member of the IEICE, the Institute of Image Information and Television Engineers (ITE), the Japan Society of Civil Engineering (JSCE), the Society of Automotive Engineering of Japan (JSAE), and the Japan Association for Human and Environmental Symbiosis (JAHES).



TAKAHIRO OGAWA (S'03–M'08–SM'18) received the B.S., M.S., and Ph.D. degrees in electronics and information engineering from Hokkaido University, Japan, in 2003, 2005, and 2007, respectively, where he is currently an Associate Professor with the Faculty of Information Science and Technology. His research interest includes multimedia signal processing and its applications. He has been an Associate Editor for the *ITE Transactions on Media Technology and Applications*. He is a member of the ACM, EURASIP, IEICE, and the Institute of Image Information and Television Engineers (ITE).

MIKI HASEYAMA (S'88–M'91–SM'06) received the B.S., M.S., and Ph.D. degrees in electronics from Hokkaido University, Japan, in 1986, 1988, and 1993, respectively. She joined the Graduate School of Information Science and Technology, Hokkaido University, as an Associate Professor, in 1994. She was a Visiting Associate Professor with Washington University, USA, from 1995 to 1996. She is currently a Professor with the Faculty of Information Science and Technology, Hokkaido University. Her research interests include image and video processing and its development in semantic analysis. She has been the Director of the International Coordination and Publicity of the IEICE. She is a member of the IEICE and the Information Processing Society of Japan (IPSJ) and a Fellow of the Institute of Image Information and Television Engineers (ITE). She has been the Vice-President of the ITE. She has been the Editor-in-Chief of the *ITE Transactions on Media Technology and Applications*.