"Does 4-4-2 exist?" – An Analytics Approach to Understand and Classify Football Team Formations in Single Match Situations

Eric Müller-Budack
eric.mueller@tib.eu
Leibniz Information Centre for Science and Technology (TIB)
Hannover, Germany

Robert Rein
r.rein@dshs-koeln.de
German Sport University Cologne
Cologne, Germany

Jonas Theiner
theiner@stud.uni-hannover.de
Leibniz Universität Hannover
Hannover, Germany

Ralph Ewerth
ralph.ewerth@tib.eu
L3S Research Center, Leibniz Universität Hannover & Leibniz Information Centre for Science and Technology (TIB)
Hannover, Germany

ABSTRACT
The chance to win a football match can be significantly increased if the right tactic is chosen and the behavior of the opposite team is well anticipated. For this reason, every professional football club employs a team of game analysts. However, at present game performance analysis is done manually and therefore highly time-consuming. Consequently, automated tools to support the analysis process are required. In this context, one of the main tasks is to summarize team formations by patterns such as 4-4-2 that can give insights into tactical instructions and patterns. In this paper, we introduce an analytics approach that automatically classifies and visualizes the team formation based on the players’ position data. We focus on single match situations instead of complete half-times or matches to provide a more detailed analysis. The novel classification approach calculates the similarity based on pre-defined templates for different tactical formations. A detailed analysis of individual match situations depending on ball possession and match segment length is provided. For this purpose, a visual summary is utilized that summarizes the team formation in a match segment. An expert annotation study is conducted that demonstrates 1) the complexity of the task and 2) the usefulness of the visualization of single situations to understand team formations. The suggested classification approach outperforms existing methods for formation classification. In particular, our approach gives insights into the shortcomings of using patterns like 4-4-2 to describe team formations.

KEYWORDS
Sports Analytics, Pattern Analysis, Formation Classification, Annotation Study

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must also be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM Reference Format:
Eric Müller-Budack, Jonas Theiner, Robert Rein, and Ralph Ewerth. 2019. "Does 4-4-2 exist?" – An Analytics Approach to Understand and Classify Football Team Formations in Single Match Situations. In 2nd International Workshop on Multimedia Content Analysis in Sports (MMSports ’19), October 25, 2019, Nice, France. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3347318.3355527

1 INTRODUCTION
The choice of the right tactic in a football match can have a decisive influence on the result and thus has a great impact on the success of professional football clubs [32]. According to Garganta [8] and Fradua et al. [7], the tactic describes how a team manages space, time and individual actions during a game. To select an appropriate tactic, detailed analyses are necessary to reveal and possibly exploit insights into opposite team’s behavior and patterns. These decisions are usually left to domain experts such as the coaching, analysts and scouting staff who observe and analyze entire football matches in order to prepare the next match. However, this process is very time-consuming which has limited its application in the past [14, 32]. For this reason, as the amount of available game performance data is steadily increasing the demand for automated analysis tools to support the scouting process is rapidly growing. Early approaches are mainly based on the interpretation of match statistics such as the distribution of the ball possession as well as shot, pass and tackle variables with the general aim of predicting successful teams [6, 15, 16, 20, 25, 27, 31]. These statistics discard most contextual information as they are usually calculated across extended game periods like half-times, whole matches or seasons. Therefore, these measures are not able to capture the increasing complexity of tactic in modern football and lack explanatory power in terms of prediction variables for game success [24]. The development of advanced tracking technologies [2, 21] by various companies [26, 34] has opened up new opportunities through the availability of accurate positional data about the players and the ball. These data allow to apply automated approaches to analyze different tactical aspects [13, 24, 35]. Referring to Rein and Memmert [32], tactics can be distinguished into team, group and individual tactics. One important aspect with respect to team tactics is the...
team formation [4, 32]. The team formation describes the spatial arrangement of the players on the pitch by dividing them into tactical groups (e.g., defenders, midfielders, and attackers). However, team sports are in general highly complex and dynamic since players are constantly switching positions and roles throughout a match. Consequently, Bialkowski et al. [4, 5] considered formation detection as a role assignment problem. First automatic approaches for formation classification assumed that the team formation is stable over a match half and thus focus on the classification of formation for whole matches or halftimes [4, 5]. More recently, spatio-temporal methods were introduced that aim to detect formation changes during the match [5, 4, 23, 38], but either evaluation was not performed on single match situations or, in case of ForVizor [38], case studies were utilized to evaluate the visual analytics system itself. Moreover, previous work [3, 38] mainly relied on clustering algorithms to automatically find the most prominent formations during a match to measure temporal differences in individual situations. Therefore, it is not possible to directly predict a numerical scheme such as 4-4-2 and the resulting clusters require supervision for classification. In contrast, Machado et al. [23] developed a visual football match analysis tool where formations are classified by a k-means clustering approach using the coordinates of the players themselves and assigning them to one of three tactical groups (defender, midfielder, attacker). Although this approach is able to directly predict formations, the clustering is solely based on one-dimensional player coordinates with respect to the field length (from goal to goal) and formations are restricted to three tactical groups.

In this paper, we present an analytics system that automatically generates a visualization summary of the formation in single match situations to capture the highly complex and dynamic characteristics of football. Thereby, a match situation is defined from gaining possession to losing ball possession. Based on the visual representation, we present a novel classification approach that calculates the similarity (of a single match formation) to ground-truth templates for twelve different popular tactical formations (some examples, e.g., 4-4-2, 4-2-3-1 are shown in Figure 1). The quality of the visualization has been evaluated by twelve domain experts that have also provided ground-truth annotations for numerical schemes of tactical formations. A detailed analysis has shown that visual summaries are helpful to quickly provide an overview of individual game situations. The inter-coder agreement of the experts is measured and provides first insights about the applicability of numerical schemes to describe team formations. It turns out that one main issue is that some tactical formations only differ in the interpretation of some roles. To address this issue, we propose a novel metric to measure the quality of the formation classification with respect to the similarity to ground-truth formation templates provided by domain experts. This also allows us to treat the recognition of formations as a multi-label or fuzzy classification task that is better applicable to real scenarios. To the best of our knowledge, this is the first work that provides a solution and detailed analysis of formation classification for single match situations.

The remainder of the work is organized as follows. Section 2 reviews related work in sports analytics with focus on formation detection in football games. Our system to create a visual formation summary and to classify the formation in single match situations is presented in Section 3. In Section 4, the experimental results based on expert annotations are discussed in detail. Section 5 summarizes the paper and outlines potential areas of future research.

2 RELATED WORK

Analytics in football or sports in general is a broad field that has recently attracted more attention. This is mostly caused by the availability of positional data commonly captured by pre-installed tracking devices in stadiums [2, 21] or provided by companies such as OptaSports [26] or STATS [34]. Since a more general overview goes beyond the scope of this work, we refer to the review of Rein and Memmert [32], that covers various aspects and challenges of automated content analysis in football. One fundamental research area is the tactical analysis of football data. We therefore focus on related work that has been introduced to find general tactical patterns as well as to explicitly classify and visualize team formations.

Many approaches have been suggested that aim to cluster and consequently find prominent movement patterns of a team [11, 12, 36, 37]. In this context, sketch-based (video) retrieval systems were introduced [1, 30, 33] that allow users to draw spatio-temporal queries on a virtual pitch to directly retrieve similar game situations. While these approaches mainly reveal individual or group tactics, another important factor with significant impact on performance is team formation. However, due to the nature of team sports, players constantly change roles and thus make formation classification a complex task. Based on hockey games, Lucey et al. [22] have shown that a role-based representation of the formation is superior compared to a representation that is solely based on the coordinates of player identities. Subsequently, Bialkowski et al. [4, 5] have introduced a role-assignment algorithm and define the formation as a set of role-aligned player positions. However, they assume that the dominant formation is stable within a match half and this coarse temporal granularity is not sufficient to describe the complex and varying tactical formations in modern football.

To solve this issue, Perl et al. presented a number of formation analysis tools [9, 10, 29] and used the neural network DyCoN [28] to determine the distribution and sequential changes in the team formation. In addition, Bialkowski et al. [3] have extended their previous systems [4, 5] and utilized the role-assignment algorithm to discover with-in match variations using two methods as a proof-of-concept: (1) clustering of role-aligned player positions and (2) calculating the distance of each frame to the mean formation of the half time. Wu et al. [38] proposed a visual analytics system called ForVizor, that distinguishes between offensive and defensive formations. Based on the role-assignment algorithm of Bialkowski et al. [4, 5] the formation changes between different match periods are visualized. But the aforementioned systems rely on the detection of the most prominent formations, e.g., by using a clustering algorithm or the average formation of the halftime in order to detect temporal changes in the formation. Therefore, their approach cannot automatically predict a numerical tactical scheme such as 4-4-2 for short match situations. Alternatively, Machado et al. [23] developed a match analysis tool and applied k-means clustering to the one dimensional y player positions (from goal to goal) itself. Each player is then assigned to one of three tactical groups to create a numeric representation. However, this approach completely neglects the x-coordinates of the players for classification.
3 TEAM FORMATION CLASSIFICATION

As the discussion of related work reveals, previous approaches for team formation analysis focused on entire half-times or matches. In contrast, we present a novel classification approach that can be applied to single match segments of arbitrary length to understand the highly complex and dynamic aspects of football and conduct an in-depth expert evaluation. In addition, such an evaluation by domain experts in terms of analyzing individual match situations with respect to the formation played was not conducted yet. Evaluation was either focused on long-term formations [3–5] or placed more emphasis on the evaluation on the tools itself [23, 38].

Our proposed system to explore football matches with respect to the team formation is introduced in this section. The definition of a team formation is provided in Section 3.1. The required input data as well as pre-processing methods are explained in Section 3.2. Based on this input information we propose an approach to create a visual formation summary (Section 3.3) that serves to classify (Section 3.4) the team formation played in single situations of a football match. An overview is illustrated in Figure 1.

3.1 Definition of Team Formation

In general, the formation describes the spatial arrangement of players within a team. Assuming that all ten players (except the goalkeeper) are on the pitch, it is defined as a set of distinct roles \( F = \{r_1, \ldots, r_{10}\} \) that are represented by their two-dimensional position \( r \in \mathbb{R}^2 \) on the football field. For simplification, these roles are often assigned to tactical groups like defenders, midfielders (defense and offensive) and attackers to generate a numeric representation. These numerical schemes, e.g., 4-2-3-1 and 4-4-2, define the tactical formation of the team and are denoted as \( F_n \) in the following.

3.2 Input Data and Pre-processing

Our system relies on two-dimensional location information of each player at each discrete timeframe \( f \). We use the normalized coordinates with respect to the width \( x \in [0, 0.7] \) and length \( y \in [0, 1.0] \) of the football pitch and preserve the aspect ratio of the field. For unification, the direction of play of the observed team is always considered from bottom to top and the position data of the players are converted accordingly. Since the aim of this work is to automatically detect formations in individual game situation, the match is first divided into temporal segments. In this context, we require information which team possesses the ball at each timeframe \( f \) and define a game segment \( S = \{f_i, \ldots, f_{i+m}\} \) containing \( m \) timeframes from gaining to losing the ball or vice versa.

3.3 Visual Formation Summary

First, a visual formation summary (VFS) using the two-dimensional position data of all frames in a game segment \( S \) is generated. Regardless of the absolute positions (e.g., while defending most players are located in the own half), the formation in terms of the relative distances between the players within a team remains the same. Therefore, we subtract the team center that is defined as the mean of all individual player positions at each timeframe for normalization. As stated in Section 3.1 the formation is defined by a set of roles. Theoretically, each player can be considered to act in one role and represented by its mean position during a match segment. However, as mentioned by previous work [3, 5], players can potentially switch roles and this approach would not accurately reflect the tactical formation. For this reason, we employ the role-assignment algorithm of Bialkowski et al. [3, 5] to detect and compensate role changes. Note, that only one iteration is applied in our system. More than one iteration did not have a great influence on the result in our experiments, which is supposedly due to the length of the sequences. Finally, the mean position \( \bar{F} \) for every role during the observed match segment is utilized to define the formation \( F = \{\bar{r}_1, \ldots, \bar{r}_{10}\} \) and to derive the VFS.

3.4 Classification of Numerical Schemes

The formation \( F \) with compensated role changes according to the previous section serves as an input to classify each game situation.
into a common numerical tactical schema like 4-4-2. We propose a novel classification approach that measures the similarity of the extracted formation \( F \) to a pre-defined set of \( t \) popular football formations \( T = \{ \hat{F}_1, \ldots, \hat{F}_t \} \). The expected player coordinates are provided by domain experts as explained in Section 4.2.

In order to enable a comparison between two formations, it is necessary to normalize the positional data of each role \( i \) by the minimum and maximum \( x \) and \( y \) coordinate within a formation \( F \):

\[
\tilde{r}_i = \frac{r_i - \min(F)}{\max(F) - \min(F)} \quad \forall r_i \in F.
\]

The formula provides also a normalization of the relative distances of the players and therefore allows for a comparison of formations with different compactness. Subsequently, a similarity matrix \( M(\hat{F}_1, \hat{F}_2) \in \mathbb{R}^{10 \times 10} \) is calculated. Since we use idealized templates for formation classification, the Euclidean squared distance is applied in this context, because it penalizes smaller distances between different roles less severely. Each entry \( m_{i,j} \) is then calculated based on the positional coordinates of each role \( \tilde{r}_i = (\tilde{x}_i, \tilde{y}_i) \) in formation \( \hat{F}_1 \) to each role \( \tilde{r}_j = (\tilde{x}_j, \tilde{y}_j) \) in formation \( \hat{F}_2 \) according to the following equation:

\[
M(\hat{F}_1, \hat{F}_2) = \max \left( 1 - \frac{||\tilde{r}_i - \tilde{r}_j||^2}{\delta}, 0 \right) \quad \forall \tilde{r}_i \in \hat{F}_1; \tilde{r}_j \in \hat{F}_2.
\]

The normalization factor \( \delta \) serves as tolerance radius. Under the assumption that a football pitch can be divided into three horizontal (left, center, right) and vertical groups (defender, midfielder, attacker), a normalization factor \( \delta = 1/3 \) means that the similarity of wingers to central player as well as, e.g., from attackers to midfielders would already become zero. In our opinion, this fits well to the task of formation classification. Please note, that we only allow similarity values in the interval range \( m(i, j) \in [0, 1] \).

To calculate the similarity of two formations, each role in formation \( \hat{F}_1 \) has to be assigned to its optimal counterpart in formation \( \hat{F}_2 \). With the constraint that each role can only be assigned once and the overall goal to maximize the similarity this results in a linear sum assignment problem that can be solved via the Hungarian algorithm [19] whose solution corresponds to:

\[
m^*_i = \begin{cases} m_{i,j}, & \text{if } \tilde{r}_i \in \hat{F}_1 \text{ is assigned to } \tilde{r}_j \in \hat{F}_2, \\ 0, & \text{otherwise}. \end{cases}
\]

Finally, the formation similarity \( \text{FSIM}(\hat{F}_1, \hat{F}_2) \) of the compared formations is defined as the sum of all elements in the similarity matrix \( M^*(\hat{F}_1, \hat{F}_2) \) normalized with respect to the number of assigned roles (in this case ten). To classify the derived formation \( F \) according to Section 3.3 of a given match segment into the numerical schema \( \hat{F} \), we calculate the similarities to a set of pre-defined templates \( T = \{ \hat{F}_1, \ldots, \hat{F}_t \} \) that contain idealized role positions for \( t \) popular football formations. This allows us to generate a ranking of the most probable numerical formation played in an individual match situation. For the final classification, the template formation with the highest similarity is selected as defined in Equation (4):

\[
\hat{F}^* = \arg\max_{F \in T} \left\{ \text{FSIM}(F, \hat{F}) \right\}
\]
Table 1: Number of annotations, mean match situation length and standard deviation ($\sigma$) of the dataset used for evaluation. A subset was annotated by two domain experts each in order to assess the quality of the annotations.

| Ball Possession | Duration | Annotations (from two experts) | Mean length [s] | $\sigma$ [s] |
|-----------------|----------|--------------------------------|-----------------|-------------|
| Own             | short    | 105 (35)                       | 6.96 (1.38)     |             |
|                 | mid      | 234 (74)                       | 13.67 (2.87)    |             |
|                 | long     | 231 (71)                       | 31.53 (12.62)   |             |
| Opponent        | short    | 91 (23)                        | 6.73 (1.30)     |             |
|                 | mid      | 232 (71)                       | 13.71 (2.95)    |             |
|                 | long     | 236 (74)                       | 32.17 (13.35)   |             |

Figure 2: Analytics tool for formation detection. The two-dimensional animation of the selected match situation (middle) is shown on the left and the resulting visual formation summary of the scene is shown on the right side.

4.2 Template Formations

As explained in Section 3.4, the classification of the played formation is performed by a comparison to a pre-defined set of templates for different tactical formations. Our domain experts were asked to provide these ground-truth templates to the best of their knowledge. But some formations like 4-4-2 contain some variations and are not completely unambiguous. Hence, multiple templates for one formation should be created. For classification we have calculated the similarity of the visual formation summary to all variations of a single formation, and used the maximum as value for the formation similarity FSIM. The templates created for all twelve formations used in the experiments are visualized in Figure 4.

4.3 Analysis of the Expert Study

Data: In the rest of our experiments, we only consider match situations where a formation was clearly or very clearly recognizable for at least one expert. This results in a total of 472 unique situations of which 207 were annotated by two experts for classification and 450 ratings for the visual formation summary.

Annotation statistics: Referring to Figure 3, the analysis of the annotated match situations has shown a bias towards some popular formations such as 4-4-2, 4-2-3-1 and 4-3-3. These results were more or less expected, since e.g. the 4-4-2 is generally widely accepted and therefore used more frequently to describe a formation compared to a 4-2-4, which however has very similar spatial properties as shown in Figure 4. More surprisingly, the majority of annotations were rated at least clearly recognizable by the experts, despite the short length of single match situations. In this context, the defensive
Table 2: Agreement of the expert annotations for detecting a formation in terms of Krippendorff’s $\alpha$ [18], the formation similarity (FSIM) and accuracy. Top-k accuracy means that the specified formation of at least one annotator is within the top-k nearest formations (according to Figure 5) of the specified formation of the other annotator.

| Ball possession          | Duration (Nr. of scenes) | Own short (13) | mid (39) | long (43) | all (95) | Opponent short (10) | mid (44) | long (58) | all (112) | Overall (207) |
|-------------------------|--------------------------|----------------|----------|-----------|----------|---------------------|----------|-----------|-----------|---------------|
| Krippendorff’s $\alpha$ [18] |                          | 0.23           | 0.25     | 0.18      | 0.22     | 0.54                | 0.20     | 0.26      | 0.27      | 0.26          |
| Formation Similarity (FSIM) |                          | 0.90           | 0.91     | 0.92      | 0.91     | 0.96                | 0.94     | 0.94      | 0.94      | 0.93          |
| Accuracy Top-1           |                          | 0.38           | 0.41     | 0.42      | 0.41     | 0.70                | 0.55     | 0.55      | 0.56      | 0.49          |
| Accuracy Top-3           |                          | 0.62           | 0.69     | 0.70      | 0.68     | 0.80                | 0.70     | 0.76      | 0.72      | 0.71          |
| Accuracy Top-5           |                          | 0.77           | 0.82     | 0.95      | 0.87     | 1.00                | 0.95     | 0.98      | 0.97      | 0.93          |

Formations were annotated with a larger confidence than offensive formations and have shown less variance (mainly 4-4-2). This effect can be explained by the increased freedom of the players during the attack to make creative plays, which are very important for scoring goals in modern football [17].

**Annotation quality:** In order to assess the quality of the provided annotation of the played formations, we first measured the inter-coder agreement with respect to Krippendorff’s $\alpha$ [18] and the top-1 accuracy. The results are reported in Table 2.

Overall, annotations for defensive formations show significantly more correlation than offensive formations. As already mentioned, this is mainly due to freedom and creativity in attacking situations that lead to more fluid formations. Most noticeably, the agreement in terms of Krippendorff’s $\alpha$ [18] and the top-1 accuracy is significantly lower than expected. However, both metrics expect the annotators to determine exactly the same formation. We believe that the annotations from domain experts still show correlations, but that the complexity and subjectivity of the task leads to different conclusions. As stated above, team formations sometimes only differ in the interpretation of very specific roles. In addition, it is possible that (1) formations are not symmetric and (2) different formations are played within a single game situation, e.g. when multiple offensive game patterns are performed during an attack.

Therefore, we propose to calculate the formation similarity (FSIM) between the templates of the annotated formations to obtain an alternative measure of the inter-coder agreement. This also enables us to measure the top-k accuracy that determines whether the specified formation of at least one annotator is within the top-k most similar formations of the specified formation of the other annotator. The similarity values of all tactical formations based on their templates are visualized in Figure 5. It is clearly visible that the top-3 accuracy is significantly better than the top-1 accuracy (Table 2) with respect to the inter-coder agreement. In addition, the formation similarities (FSIM) between the annotated formations are comparably high, especially if the values from Figure 5 are taken into account. From our point of view, this indicates that the annotations of experts do indeed show a high correlation, at least if the recognition of formations is treated as a multi-label task where more than one answer can be considered as correct.

### 4.4 Evaluation of the VFS

**Metrics:** As explained in Section 4.1.2, the domain experts were asked to rate the usefulness of the extracted visual formation summary (VFS) of a team in a given match situation. In addition, we quantified the formation similarity (FSIM) of the VFS to the template of the annotated formation. The results are reported in Table 3.

**Results:** Overall, the VFS’s were mainly rated positive and only in 16% of the situations the annotator did not see correlations to the two-dimensional schematic visual representations. This demonstrates that the VFS indeed provides a good overview in the majority of the cases. Particularly, in situations with opposing ball possession, the VFS can quickly give insights into the tactical defensive formation and therefore simplifies the analysts’ process. The same conclusions can be drawn with respect to the obtained formation similarity of the extracted VFS to the templates of the annotated formation. Similarities around 0.75 and 0.80 are achieved in the two cases of own and opposing team ball possession, respectively. Although these values are comparatively lower than the template similarities in Table 5, we believe that the results indicate a satisfying system quality. The template similarities are calculated based
Table 3: Evaluation results of the visual formation summary (VFS) in terms of expert ratings and the formation similarity FSIM to the annotated formation of the expert.

| Ball possession | Duration (#Scenes) | Rating | FSIM |
|-----------------|--------------------|--------|------|
| Own             | short (23)         | 0.26   | 0.50 | 0.43 | 0.77 |
|                 | mid (87)           | 0.22   | 0.33 | 0.45 | 0.76 |
|                 | long (96)          | 0.15   | 0.38 | 0.48 | 0.76 |
|                 | all (206)          | 0.19   | 0.34 | 0.46 | 0.76 |
| Opponent        | short (24))        | 0.08   | 0.33 | 0.58 | 0.78 |
|                 | mid (98)           | 0.14   | 0.26 | 0.60 | 0.80 |
|                 | long (122)         | 0.16   | 0.30 | 0.55 | 0.81 |
|                 | all (244)          | 0.14   | 0.28 | 0.57 | 0.80 |
| Overall (450)   |                   | 0.16   | 0.31 | 0.52 | 0.78 |

Analysis and discussion of the results: Although the results are improved compared to previous baselines, in particular the top-1 accuracy is rather low. To analyze possible problems, quantitative as well as qualitative results are provided in Table 5 and Figure 6. The quantitative results again show that the task of formation classification becomes easier for defense situations and for situations with increasing length. Additionally, results improve significantly when considering the top-k, particularly for $k > 2$, similar formations. Referring to the qualitative results in Figure 6, this can be mainly explained by the problems described below. As previously stated, formations can be very similar and their classification often depends on the interpretation of specific roles. In particular, wing backs in formations with four defenders are moving up on the pitch to get involved in the attack or pro-actively defend in pressing situations. This is clearly visible in scenes 1, 4 and 6. While the domain expert in scene 1 considers them as midfielders, they are mostly perceived as defenders in similar situations. However, simultaneously the defensive midfielder could move back to form a three-man formation with both center backs. For this reason, e.g., a 4-4-2 or 4-3-3 is often classified as a 3-4-3 or 3-5-2 by our system, as depicted in the confusion matrix (Figure 6 right). Similarly, offensive wingers or attacking midfielders can be either interpreted as midfielders or strikers. Referring to Figure 3, the experts lean towards more popular formations such as 4-4-2 during their annotations, while the visual formation summary suggests a 4-2-4 instead with respect to the pre-defined templates (scene 5). Admittedly, the experts have classified the formation based on the two-dimensional graphical animation representation and also had context from preceding situations which can have influence on the rating and allows for other conclusions. But in many cases the annotators even considered the VFS as good, which shows that the mistakes are often connected to the subjective interpretation of roles instead of the similarity to idealized templates. Overall, the analysis suggests that formation approach. The results are also much better than always predicting 4-4-2 in terms of micro-accuracy and random guessing for both metrics. Furthermore, we could confirm that the role assignment algorithm improves team formation classification.

4.5 Evaluation of the Formation Classification

Metrics: Referring to Figure 3, the data set contains a large bias towards some formations. Therefore, we report micro-accuracy alongside macro-accuracy as it allows us to study system’s classification performance while considering each class to be equally important. Additionally, the top-k accuracy is reported. In this context, the VFS of the observed scene is compared to all available template formations to generate a ranking with respect to the formation similarities (FSIM). Please note, that some match situations were analyzed by two experts and their annotations can differ. We assume that both annotations are valid and use the annotated formation which has a higher similarity to the VFS as reference.

Baseline: As discussed in the related work section, previous work [3, 38] apply clustering approaches to find the most prominent formations in a match in order to measure formation changes. These approaches are not capable to automatically classify the formation and are therefore not suitable for comparisons. For this reason, we can only compare our proposed classification approach to Machado et al. [23]’s system. However, their solution relies on a k-means clustering of y-coordinates and can only predict a predefined amount of, in this case $k = 3$, tactical groups. In addition, it could predict unrealistic formations such as 2-7-1. This is a systematic drawback and the expert annotations shown in Figure 3 indicate that other formations are labelled very rarely.

Comparison to baseline approaches: To enable a comparison, we reduce the number of groups in the annotated formations as well as predictions from four to three by assigning the most similar formation with three tactical groups according to the similarity values in Figure 5. The annotated 4-1-4-1 and 4-2-3-1 formations become a 4-5-1 and the 4-3-2-1 is converted to a 4-3-3, yielding a total number of nine different classes. Machado et al. [23]’s classification approach based on the clustering of y-coordinates is also applied to the visual formation summaries and thus to the same input data as our system. In addition, we investigate the impact of the role assignment algorithm. The results are reported in Table 4 and clearly show that our classification approach is superior to Machado et al. [23]’s approach. The results are also much better than always predicting 4-4-2 in terms of micro-accuracy and random guessing for both metrics. Furthermore, we could confirm that the role assignment algorithm improves team formation classification.
Table 5: Evaluation results of the classification approach in terms of the micro and macro top-k accuracy. Note, that only match situations are considered in which at least one expert could clearly or very clearly identify the formation.

| Ball Possession Duration (Nr. of scenes) | Own | | | Opponent | Overall (472) |
|---|---|---|---|---|---|
| | short (31) | mid (94) | long (97) | all (222) | short (30) | mid (100) | long (120) | all (250) |---|
| Macro Accuracy | Top-1 | 0.13 | 0.09 | 0.13 | 0.11 | 0.30 | 0.24 | 0.30 | 0.28 | 0.20 |
| | Top-2 | 0.19 | 0.20 | 0.29 | 0.24 | 0.40 | 0.47 | 0.45 | 0.45 | 0.35 |
| | Top-3 | 0.45 | 0.33 | 0.52 | 0.43 | 0.53 | 0.63 | 0.60 | 0.60 | 0.52 |
| | Top-4 | 0.55 | 0.51 | 0.66 | 0.58 | 0.77 | 0.71 | 0.73 | 0.73 | 0.66 |
| | Top-5 | 0.61 | 0.63 | 0.73 | 0.67 | 0.80 | 0.80 | 0.81 | 0.80 | 0.74 |
| Micro Accuracy | Top-1 | 0.20 | 0.09 | 0.15 | 0.10 | 0.19 | 0.28 | 0.29 | 0.23 | 0.20 |
| | Top-2 | 0.24 | 0.16 | 0.25 | 0.16 | 0.25 | 0.38 | 0.36 | 0.30 | 0.27 |
| | Top-3 | 0.43 | 0.23 | 0.40 | 0.25 | 0.33 | 0.58 | 0.46 | 0.40 | 0.39 |
| | Top-4 | 0.46 | 0.34 | 0.49 | 0.33 | 0.58 | 0.61 | 0.53 | 0.53 | 0.48 |
| | Top-5 | 0.49 | 0.48 | 0.57 | 0.46 | 0.63 | 0.64 | 0.58 | 0.56 | 0.53 |

Figure 6: Qualitative results (left) of the proposed analytics system for six individual scenes with the respective top-3 tactical group assignment (emphasized in different colors) after comparing with the templates in Figure 4 as well as the confusion matrix (right) for the predictions of all 472 match situations in percent.

Classification should be considered as a multi-label or fuzzy classification task, where more than one answer could be correct. For this reason, we believe that a VFS often provide valuable insights into the tactical formations. The formations’ similarities to templates of popular formation, however, can help to monitor tactical changes as well as to retrieve situations to show specific formations.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we have presented an analytics approach for the visualization and classification of tactical formations in single situations of football matches. A novel classification approach has been proposed exploiting a set of ground truth templates that contains idealized player positions for twelve popular team formations. A detailed analysis of an expert annotation study was conducted to provide results for defensive and offensive formation classification in match situations with various length. The study has clearly demonstrated the complexity of the task, particularly for offensive formations, since even annotations from domain experts differ due to the subjectivity in interpreting roles of similar formation schemes such as 4-2-3-1 and 4-3-3. For this reason, we have suggested a novel measurement to quantify the results for formation classification and visualization based on the similarity to pre-defined formation templates. The results demonstrated that our visual formation summary already provides valuable information and is capable to summarize individual scenes in football matches. In addition, we have shown the superiority of our classification approach compared to the current state of the art.

In the future, we plan to extend the current analytics system with other valuable tactical indicators such as the variance and movements of the players. Additionally, we aim to exploit information of previous and subsequent situations to improve the system performance. Our current approach explicitly aimed for a solution that does not require any training data for classification. However, due to the increasing amount of position data, whether synthetic or real, deep learning approaches could become applicable to find more sophisticated solutions for the classification of team formations.
