Spatio-Temporal Analysis for Sports Play Retrieval

Zheng Wang*, Xia Yunwen*, Ji Yitong*, Jiang Yuan*
*School of Computer Science and Engineering, Nanyang Technological University, Singapore

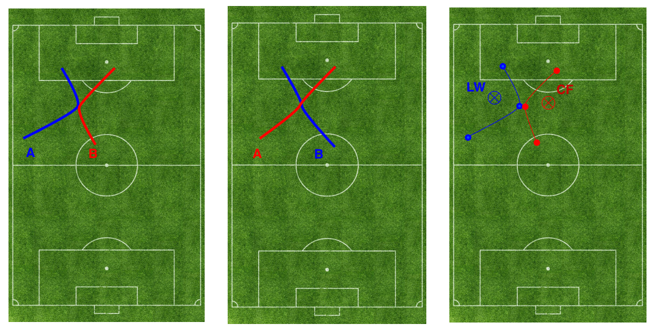
***Abstract*—Nowadays, many commercial systems (e.g., SportVU) have been widely deployed to track players' movement during a sports game. The data collected by these tracking systems can describe both spatial and temporal features of a game, and thus it is often termed as spatiotemporal sports data. Meanwhile, the data has also attracted much research attention and facilitated plenty of works further such as formation detection, tactics discovery and similar play retrieval. In this essay, we focus on one of them called sports play retrieval, which is usually considered as a fundamental operator in many sports applications (e.g., recommendation). We comprehensively review existing research studies in sports play retrieval and some closely related analytical tasks. Besides, we conclude four widely used algorithmic approaches in terms of retrieval techniques (i.e., Keyword Search, Trajectory Similarity, Role Detection and Deep Learning) in this topic. Finally, we outline the essential evaluation methods that a quality retrieval system should possess in order to maximize user experience.**

I. Introduction

With the wide-spread of sports tracking systems, the topic of sports analysis has gained a lot of attention [1], [2], [3], [4] in recent years. For example, SportVU developed by STATS LLC generates tracking data for every player, ball, and referee at a frequency of 10Hz in a soccer game. Meanwhile, these huge amounts of tracking data record many event details such as passes, shots and dribbling, and the interest in effective and efficient sports information retrieval has correspondingly grown as well. In many sports domain, a play is a basic unit, which describes the trajectories of moving multiple agents (e.g., multiple players in a sports game) over time. Additionally, these collected data can fully describe the process in a sports game (i.e., it embeds both spatial and temporal features of a game). Hence, it is usually termed as spatiotemporal sports data.

Quantifying the similarity between spatiotemporal sports data (e.g., multi-agent trajectories or plays) is a fundamental research problem and is helpful for a variety of real-world applications. For example, some emerging sports applications such as ESPN and Team Stream allow users to search for relevant videos or recommend similar game plays. From a sports coach standpoint, retrieving similar game plays would be highly beneficial to further improve team tactics when preparing for an upcoming match as discussed in [2].

However, unlike other simple data types such as geometric points, where the similarity definition is straightforward (i.e., Euclidean distance), measuring similarity between plays needs to be carefully defined in order to reflect the true underlying distance. More specifically, that is how to meaningfully model play similarity. Conventional methods for sports play retrieval are based on querying some "keywords", which however requires the data to be annotated with keywords and users to have necessary prior knowledge on keywords. We call these methods Keyword-based Retrieval (KbR).



(a) (b) (c)

Fig. 1: Examples in modeling play similarity.

In terms of spatiotemporal sports data (i.e., play data), a straightforward method is to consider an agent-to-agent trajectory comparison based on pairwise point-matching methods such as DTW [5] and compute the similarity through finding an optimal matching. On the other hand, it can also take the multi-agent trajectories as a whole, and study a novel similarity measure such as EMDT [6]. It borrowed the idea of the Earth Mover's Distance to define the similarity over the trajectory sets. We call this kind of methods is Similarity-based Retrieval (SbR).

However, let us consider a real and frequently occurring example. Figure 1(a) shows two players' trajectories in a play. Figure 1(b) shows another two players' trajectories. The trajectory A in Figure 1(a) is not similar to the trajectory A or the trajectory B in Figure 1(b). Similarly, it is the same for the trajectory B in Figure 1(a). Although these two plays (a) and (b) are very similar if you look at the whole, the agent-to-agent method will give a large matching cost and miss the better results. Recent studies [7], [1], [4] proposed a role-based fast alignment method in a play. As shown in Figure 1(c), these methods divide the trajectory into many segments by running the sliding window and assign each player a unique role at each segment. Overall, These methods are proposed for sports play retrieval to overcome the exhaustive comparing problem of an agent-to-agent method by using a role-based representation to enable fast alignment of trajectories. Additionally, effective templating and hashing techniques are employed to support users' queries at interactive speeds. We call them as Alignment-based Retrieval (AbR).

However, the inherent difficulty of using the raw multi-agent data is that of the misalignment due to the constant changing of player positions or team formation over the course of a play. Additionally, the matching weights of the roles

TABLE I: Works on Sports Play Retrieval.

Method Type	Research Papers
Keyword-based Retrieval (KbR)	[12], [13], [14], [15], [16], [17], [18]
Similarity-based Retrieval (SbR)	[6], [5], [19]
Alignment-based Retrieval (AbR)	[4], [1], [20], [7]
Learning-based Retrieval (LbR)	[8], [9]

should also be different and difficult to determine because it is more intuitive for people to pay more attention to the trajectories, which are near the ball. To handle the issue, some deep learning techniques such as representation learning were proposed to measure play similarity [8], [9]. For example, play2vec is a state-of-the-art method for the problem of sports play retrieval, which is inspired by the success of modeling word similarity, and proposed to view a play as a sentence and split each play into small non-overlapping pieces as words that compose the sentence, it extended the Skip-Gram with Negative Sampling (SGNS) model [10], [11] to achieve the distributed representations for each piece. Additionally, it further proposed a denoising sequential encoder-decoder model based on the maximize the probability of recovering the clean tokens from the corrupted initial inputs to handle sampling errors and/or measurement errors such as noise, which is inevitable in the data, namely the errors due to the sampling nature and/or those due to the measurement of devices. Another benefit for these Leaning-based Retrieval (LbR) methods is the efficiency. From the perspective of a retrieval system design, it would highly demand to further improve retrieval speeds. However, the SbR methods such as the agent-to-agent, which is exhaustively comparing all trajectories in a play. This is very time consuming and impractical when the amount of data is extremely large. Moreover, although some AbR methods such as role-based can quickly align trajectories, they still cannot reduce the $O(n^2)$ (n is the mean length of the trajectories in sports play) time complexity of the pairwise point-matching methods and incur many sports mining tasks limited to moderately-sized datasets. Taking the DTW distance as an example, it may take more than 3 hours to compute each query for merely 7000 plays on a high-end server [9]. However, LbR methods such as play2vec takes a linear time to compute the similarity and can be preprocessed offline, while all the existing approaches require at least $O(n^2)$ time. In practice, it can be observed at least an order of magnitude speedup over the other methods, while yielding more user-satisfactory results.

Being motivated by the importance of the topic and plenty of research efforts, we need a systematic survey that concludes the algorithmic approaches to retrieval spatiotemporal sports data, as well as the tasks and problems that have been identified in this research area.

Overall, the essay contains the following sections, and related literature are summarised in Table I.

- In Section II, we describe two primary data using in sports play retrieval, namely text (i.e., some descriptions input by users or some keywords converted from sports video) and play data (i.e, spatiotemporal sports data, a set of multiple trajectories). We will introduce the main properties of these data, and some necessary background for readers' reference.
- Section III gives the problem definition of sports play retrieval, and four types of query language for interacting between users and retrieval systems.
- We conclude existing methods in terms of KbR, SbR, AbR and LbR as we discussed above in Section IV.
- In Section V, we introduce three evaluation methods that a quality retrieval system should consider and possess, namely robustness, efficiency and effectiveness.

We illustrate the overall framework of sports play retrieval in Figure 2. Finally, we conclude this essay and discuss some potential research directions in Section VI.

II. Sports Data

In this section, we provide more details about two types of data (i.e. Textual and Spatio-temporal data) for information retrieval in sports domain. Textual data is mainly used in KbR while spatio-temporal data is used in SbR, AbR and LbR, as described in Section IV.

A. Textual Data

For information retrieval in sports domain, textual data is widely used. Keywords are used for event detection in [14], [15], [16], [17], which enables better description of sports video for retrieval system because important events are highlighted for query matching. These keywords are drafted by professionbals in sports domain. Other than that, web-casting text is highly related to sports play, thus it is usually used as side information in information retrieval. It is widely available online in many sports websites like BBC [1] and ESPN [2], and easily accessible. Important events of sports games are highlighted in web-casting text, making it easier to extract useful information than from noisy spatio-temporal data. [18] aligns web-casting text with sports video for better annotation and indexing of sports video. This process enables optimized match between keywords query input provide by users and sports video.

Textual data can be useful in sports play retrieval and it requires keywords filtering and selection. Therefore, more manual effort and professional knowledge are required.

B. Spatio-temporal Data

With adoption of new technology to capture spatio-temporal data from team sports, there is increasing number of datasets for research. Most of these datasets are from football and basketball matches. Spatio-temporal data can be categorized into two types. The first type is *Object Trajectories* which captures the movement of players or the ball. The second type is *Event Logs* that records where and when the events take place. These events could be passes, shots at goal and etc. These two types of data can be complementary for analysis of sports matches because it provides different aspects of play. For example, location where *shots at goal* take place could be useful in measuring the qualitative ratings of the shots but this piece of information is not available in *Object Trajectories*. More details of the above-mentioned types of dataset will be discussed in this section.

[1] http://news.bbc.co.uk/sport2/hi/football/teams/
[2] http://sports.espn.go.com

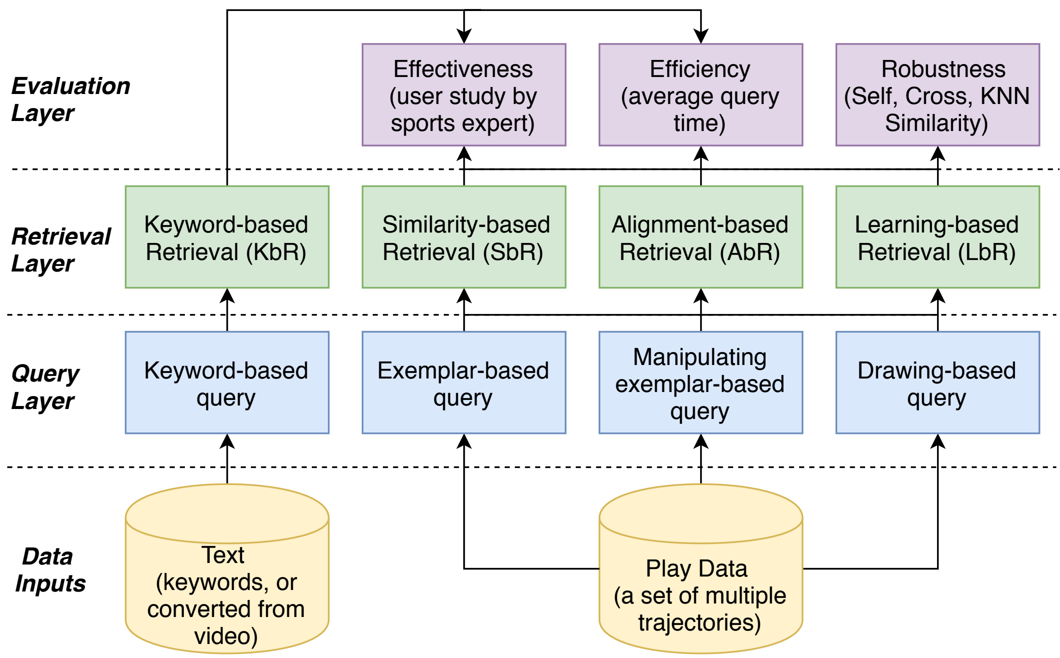


Fig. 2: Overview of Sports Play Retrieval, with each layer corresponding to a section of this essay (i.e., data inputs in Section II, query layer in Section III, retrieval layer in Section IV and evaluation layer in Section V). The solid arrows indicate the retrieval process, and the techniques described in a lower layer can be considered as input to the techniques described in a higher layer.

1) Object Trajectories: Generally, object trajectories data contains timestamped sequence of location points in a match. Optical and device tracking systems are used to record such spatio-temporal data. Recorded data will be further processed into data for research analysis. For example, [21] uses one or more cameras to capture balls' and players' motions. Trajectories of balls and players are then extracted from videos for analysis. Another popular public dataset is Prozone [3]. [22] uses entire season's data from Prozone for automatic formation detection. Prozone tracking data is also used for prediction of team identity. [22] identifies team's playing style using tracking data from Prozone.

2) Event Logs: Event log data comprises of player events like passes and shots as well as technical events such as time-outs. This type of data can be semantically richer than object trajectories data because they have more details like types of events as well as players involved. Prozone contains ball event information too. [23] utilizes event logs to measure shot quality of players. Other than just event location and time, [24] makes use of NBA tracking data which contains more event related information such as shot distances and shot angles. [24] demonstrates that these information is useful in evaluation of shooting skill and shot quality.

III. Tasks in Sports Play Retrieval

A. Problem Statement

With the development of camera, GPS and tracking system, the data of sports play, such as football game or basketball game, has shown an explosive growth trend and effective data searching has got more attention. Sports play retrieval is a process to find plays from the huge database effectively and accurately according to the user query. A play means a fragment of the sports game and its duration varies from seconds to minutes. The searching result can help coaches to prepare for match or help sports fans to find the game fragment they want to watch.

A sports play consists of the movements of the ball and all players involved in that game. The movements of these objects can be formulated as a location sequence that order by time and this sequence is defined as trajectory. The individual trajectory can be described as a vector $(x_1, y_1, x_2, y_2, ..., x_t, y_t, ..., x_F, y_F)$, (x_t, y_t) is the location at time t and F is the duration of this play. A play is composed of all trajectories in that game.

B. Query Language

The current research on sports play retrieval mainly has two types of query language. One is the text-based query [12], [13], [14], [15], [16], [17], [18] and the another is vision-based query [4], [1], [20], [8], [9].

Text-based query require user to input some keywords of the sports play and then the system will return the plays related to these keywords. Such paradigm has two main problems, one is the challenge of accurately annotating spatiotemporal data, and the other is the requirement for users. The former has been studied in [14], [15], [16] and technology is getting more and more mature. However, most current technology still has a drawback that one keyword is often related to several plays and lacks specificity. User needs to look through a large collection of plays to find the interested one. The latter problem is also

[3] http://www.prozonesports.com/

difficult to solve, and the text-based retrieval system always require the user to know the professional vocabulary before searching.

In order to create more effective and intuitive query, [1] propose a novel query paradigm called 'ChalkBoarding' and [4], [20], [8], [9] do the research in a similar way. 'ChalkBoarding' provides three different input types, *Exemplar-based query*, and *Manipulating exemplar-based queries* and *Drawing-based query*. *Exemplar-based query* allows user to search similar plays based on a fragment of existing game displayed in the window. *Manipulating exemplar-based query* lets user can adjust the trajectories of existing plays shown in the window and then use the revised one as query. For example, user can remove trajectory of irrelevant players based on his interest. *Drawing-based query* means that user can draw the trajectory completely by himself.

C. Fast Indexing

Fast indexing is also an important research area in current information retrieval system, because it is related to the fast search in the database. Most papers choose to use hash table to solve this problem. [1], [20] use the locality sensitive hashing(LSH) [25], [26] when creating the index.

D. Related Tasks

Apart from the research about sports play retrieval, there are several related tasks using spatiotemporal data and motivating the development of sports play retrieval. [27], [4] focus on the team formation detection, which lead to the role-based alignment in similar play retrieval. [2], [28] discuss the movement pattern mining which can reduce the manpower for reviewing video and scouting matches. [29], [3] study the score prediction, which is helpful to decision making in team sports.

IV. Methodology

A. Keyword-based Retrieval (KbR)

Conventional methods for sports related retrieval have been focusing on the utilization of keywords. Keywords have been used in query for sports related materials. [12] analyses keywords query logs to understand sports information seeking behaviours. The author explains the association between sport and query terms, uncovering associative subjects and topics of a sport. Through query processing, tokenization, stop word filtering, data filtering and co-occurrence detection, this paper proposes a method that would identify appropriate search keywords, which optimizes query process. [13] further improves search accuracy by refining user's initial query via Query Expansion techniques. This paper focuses on sports domain and constructs ontology for sports. Sports related information can be organized in structured way. Ontology for sports assists the expansion of keyword query by incorporation of relevant information. This process increases search accuracy.

In addition, keyword has been used in retrieval of information in sports domain. It has been widely used for various applications such as sports event detection and sports summarization. [14] converts sports video into keyword sequences and models the temporal pattern of these sequences using hidden markov model (HMM) classifier. A long sport game can be divided into various parts including certain highlights like comer kick, free kick and etc. In the first layer, low-level features such as frequency are extracted from visual and audio streams of video. The second layer utilizes Support Vector Machine (SVM) classifiers to group features into keywords like "Whistle", "Silence" and etc. A sequence of keywords for the video is then created and HMM is applied in the last layer to classify these sequences for event detection. [15] proposes a similar method as [14]. Differences lie in the low-level features they used for classification into audio and visual keywords. [16] also utilizes audio keywords for event detection. It is similar to [14]. For example, low-level audio features like zero-crossing rate is classified into keywords using SVM classifier. Instead of using HMM classifier to model temporal pattern in the high-level analysis, sports game specific rules are used to detect event from keywords. Another paper [17] makes use of both visual and audio keywords for event detection. Similar to [16], [17] drafts sport game specific rules to detect event using the visual and audio keywords extracted. Apart from obtaining keywords from sports video, keywords can be obtained from external sources. [18] proposes a generic framework by combining alignment of web-casting text and broadcast sports video for annotation, event detection, summarization and personalized retrieval of sports games. Keywords are identified from web-casting text. Events can be detected by alignment of keywords and videos. Sports videos are then annotated and indexed to construct a sports video retrieval system.

B. Similarity-based Retrieval (SbR)

Similarity-based retrieval is a process of finding those plays from a database that are similar to a query play, where a play corresponds to a fragment of a game and has its duration varying from seconds to minutes. It is widely used in some emerging sports applications such as ESPN and Team Stream to recommend similar plays to sports fans. Besides, it could help sports club managers and coaches to improve team tactics when preparing for an upcoming match [2]. The core problem of similar play retrieval is measuring the similarity between two plays, which is non-trivial because each play involves multiple trajectories. Existing solutions for this problem all adopt a two-step approach: (1) it aligns the trajectories in one play to those in the other; and (2) it computes the similarity between each pair of trajectories that have been aligned to each other using some trajectory similarity metrics such as the Dynamic Time Warping (DTW) [5] and then sums up all similarities to be one between the two plays [1]. Two strategies have been considered for the alignment step [1]. The first one is an enumeration-based strategy which (conceptually) considers all possible alignments and picks the one with the best similarity score. In practice, the method with this alignment strategy could be implemented by finding the optimal matching between the two sets of trajectories using algorithms such as the Hungarian algorithm, where the weight between a pair of trajectories is set to be the similarity between them. The second one is a role-based strategy which first detects the role of each player, which is hidden, and then aligns the trajectories of two players who share their roles. Here, the role of a player detected may have semantics such as center forward, midfield, etc. in a soccer game and could change from time to time throughout the game.

The problem of measuring the similarity between trajectories (time series in general) has been studied extensively. DTW [5] is the first attempt towards solving the local time shift issue for computing trajectory similarity. The idea is to make use of the given distance measures to map sequences into points in k-d space, and then build an index structure and to determine a cheaply computed lower bound on the original distance function, that can be used as a filter to discard non-qualifying sequences quickly. Frechet distance [19] is a classical similarity measure that treats each trajectory as a spatial curve and takes into account the location and order of the sampling points. ERP [30] and EDR [31] are proposed to further improve the ability to capture spatial semantics in trajectories. Nevertheless, these methods are mainly based on alignment of matching sample points, and thus they are inherently sensitive to noise and varying sampling rates which exist commonly in trajectory data. To address this issue, Su et al. [26] pioneer a systematic approach to trajectory calibration that is a process to transform a heterogeneous trajectory dataset to one with (almost) unified sampling strategies and then propose an anchor-based calibration system that aligns trajectories to a set of fixed locations. Ranu et al. [32] formulate a robust distance function called EDwP to match trajectories under inconsistent and variable sampling rates. These similarity measures are usually based on the dynamic programming technique to identify the optimal alignment which leads to $O(n^2)$ computation complexity, where n is the length of the trajectories. More recently, Li et al. [33] propose the first deep learning approach to learn representations of trajectories in the form of vectors and then measure the similarity between two trajectories as the Euclidean distance between their corresponding vectors, which is robust to low data quality. Yao et al. [34] employ deep metric learning to accelerate trajectory similarity computation, which is generic to accommodate any existing trajectory measure and fast to compute the similarity of a given trajectory pair in linear time and elastic to collaborate with all spatial-based trajectory indexing methods to reduce the search space. The main difference between [9] and these studies is that [9] is on plays (which correspond to sets of multiple trajectories) while these existing ones are on single trajectory. Another related study is one studying trajectory set similarity on road networks by He et al. [6], in which the idea of the Earth Mover's Distance (EMD) is leveraged to capture both spatial and temporal characteristics of trajectories.

C. Alignment-based Retrieval (AbR)

The sports tracking data records trajectory sequences of the players and the ball, and a crucial challenge of retrieving such data is how to process multiply trajectory sequences in one play. Therefore, alignment of these trajectories between two plays has received more and more attention in sports play retrieval since the development of spatiotemporal data.

The basic solution of this problem is to use an enumeration-based method, which exhaustively comparing all trajectories in two plays. This strategy would lead to the optimal alignment result, but it has great complexity and low efficiency. Several strategy, including role-based alignment and tree-based alignment, has been proposed to address this problem effectively.

The role-based alignment is inspired from the role-based representation, which analyze team behavior and individual behavior of each players. Following [27], [4], [1] calculate the formation template and assign each player to a role with the Hungarian algorithm [35]. The formation template describes the spatial arrangement, such as point guard, shooting guard, small forward, power forward and center, of all players. The formation is built through a manual annotated codebook in [27] and [4] use an approach like k-means clustering to find the formation automatically and directly from data. There are also several studies [36], [37], [38] about role-based alignment, which has not been utilized in sports play retrieval, but helpful to retrieve information in a play. Such role-based alignment can find a near-optimal alignment result more effectively and it is robust to substitutions between different players, regardless of the individual identity. However, it has an unreasonable assumption that the observed behavior is linear which consists of single state. To correct this assumption, [20] propose a tree-based alignment which learn a separate template for each state of on behaviors. The tree-based approach will first pick up the coarse template like the role-based alignment and then partition the data into various states to find a better alignment.

These alignment strategy are often combined with some relevance estimation metrics, such as Euclidean distance, Frechet distance, edit distance, longest common subsequence (LCSS) and dynamic time warping (DTW) [5], to get the similarity between two plays. According to the experimental result in [1], DTW shows best accuracy in the experiment and Euclidean distance is the most effective one for sports play retrieval among these metrics.

D. Learning-based Retrieval (LbR)

There are plenty of research studies using neural networks based methods in sports data analytics such as sports play prediction [29], [39], [40], [41], [42] or classification [43]. Specifically, Yue et al. [29] focus on basketball gameplay and presented to predict dynamic in-game events when given the current game conditions such as shooting or passing in a game. They proposed a fine-grained spatial model by adapting a discriminative learning approach (i.e., Conditional Random Fields [44]) and combining the techniques in matrix factorization.

In recent years, many deep learning-based methods towards retrieving similar sports plays have been received much research attention. Di et al. [8] proposed to apply learning to rank in sports play retrieval for processing industry-sized data. It built on the previous sports information retrieval system (i.e., chalkboarding) and extracted features from sports plays by using CNNs on the visual representations of trajectories. In addition, it used the extracted features together with some other features to learn a rankSVM model for serving users with specific preferences (conveyed with click-through data). Very recently, wang et al. [9] propose a data-driven approach to measure the similarity between two sports plays based on a deep representation learning model called play2vec. It achieves state-of-the-art performances (i.e., effectiveness and efficiency) compared with existing solutions. Additionally, it is robust against noise, which is inevitable in the spatiotemporal sports data, namely the noises due to the sampling nature and those due to the measurement of devices such as GPS devices and/or cameras.

V. Evaluation Method

A. Robustness Evaluation

A important challenge for sports play retrieval is robustness (against sampling errors such as noise and measurement errors such as non-uniform sampling rates). As we discussed before, the spatiotemporal sports data is collected by sports tracking devices such as GPS or sensor. The two types of errors are inevitable in these data, especially when the players move at high speed. However, some of existing solutions for sports play retrieval mainly focus on forming a pairwise matching of sample points of two players' (or roles') trajectories such as chalkboarding [1]. As studied in [32], [45], [33], the pairwise point-matching methods may falsely match the two trajectories even when the sequences of sampling points represent the same route. Thus, play2vec [9] considered robustness test for this topic in terms of three aspects, namely self-similarity comparison, cross-similarity comparison and KNN-similarity comparison.

1) Self-similarity Comparison: In self-similarity evaluation, the experiment first will choose some of plays from database to form the query set (denoted as Q) and some of plays as the target database (denoted as D) from the whole database. Then, for each play $P \in Q$, it will create two sub-plays by randomly sampling partial points for each trajectory in the play, denoted as P_a and P_b, and use them to construct two datasets $Q_a = P_a$ and $Q_b = P_b$. Similarly, it can get D_a and D_b from the target database D. Then for each query $P_a \in Q_a$, they compute the rank of P_b in the database $Q_b \cup D_b$ using different methods of sports play retrieval. Ideally, P_b should be ranked at the top since P_a and P_b are generated from the same play P. Furthermore, it can be used to evaluate the robustness of different retrieval approaches to noise and non-uniform sampling. For example, it can corrupt each trajectory of each play in both Q_a and $Q_b \cup D_b$ by randomly sampling a fraction of the points (corresponding to noise rate) and then for each sample point it distorts the coordinate values by adding Gaussian noises with a standard normal distribution. Similarly, it can drop a fraction of points (corresponding to non-uniform sampling rate) from each trajectory of each play in both Q_a and $Q_b \cup D_b$.

2) Cross-similarity Comparison: The targets of cross-similarity comparison is to preserve the distance between two sports plays regardless of the sampling rate or noise interference. namely Cross Distance Deviation (CDD) as follows.

$$CDD = \frac{|S(P_{a'}(r), P_{b'}(r)) - S(P_a, P_b)|}{S(P_a, P_b)}, \tag{1}$$

where $S(\cdot,\cdot)$ is a similarity measure corresponding to a specific approach for sports play retrieval; P_a and P_b are two original sports plays; $P_{a'}(r)$ and $P_{b'}(r)$ are their variants that are obtained by randomly dropping points (or adding noise) with rate r. Thus, a small CDD value indicates that an algorithm is robust and is able to preserve the original distance well.

3) KNN-similarity Comparison: Generally, the problem of sports play retrieval lacks the ground-truth. To circumvent the issue, KNN-similarity Comparison is a good evaluation method. Firstly, it needs to randomly select some of plays as the query set and some of plays as the target database from the whole database, and for each query, it conducts play retrieval to find its Top-K plays as the ground-truth of each method; Then, it needs to corrupt each play in the target database by randomly dropping points or adding noise, and retrieve the Top-K plays from the corrupted database again. Finally, it compares the retrieved Top-K plays against the ground-truth to compute the precision (i.e., the proportion of true Top-K plays among the retrieved Top-K plays).

B. Efficiency Evaluation

Efficiency evaluation is another important metric for sports retrieval systems, and it is almost considered in all existing works. The targets of this evaluation is to measure the average retrieval time of computing the similarity between a query play and the plays stored in the target database. In general, some advance indexing techniques are needed for the retrieval. For example, in chalkboarding [1], it improves retrieval speed by using a hash-table indexing [46]. In terms of play2vec [9], it takes linear time to retrieve similar plays by embedding a play into a vector. In addition, it supports preprocessing operation (i.e., encoding the plays into vectors offline); then, its online complexity becomes $O(|v|)$, where v denotes the length of the vector, which makes it scale well on large databases.

C. Effectiveness Evaluation

User study is a widely used method for effectiveness evaluation in sports play retrieval. The goal of user study for this problem is to compare the search quality of rankings generated by different retrieval methods. It generally needs to recruit volunteers with strong related background to annotate the relevance of retrieved plays, and there are two necessary points need to be noted in the study process. (a) It needs to get every participant understand the interface (e.g., images, videos) that you shown. (b) This is a blind test, which means volunteers do not know which result is from which method, and it is recommended to use Team-Draft Interleaving method [47].

VI. Conclusion and Future Research Directions

The topic of sports play retrieval has inspired many research efforts and applications in both academia and industry.

In this essay, we comprehensively review existing research problems and tasks that have been identified in this field, and conclude the research works for retrieving spatiotemporal sports data in terms of algorithmic approaches (i.e., Keyword-based Retrieval, Similarity-based Retrieval, Alignment-based Retrieval and learning-based Retrieval). Besides, we summary three existing evaluation methods (i.e., robustness evaluation, effectiveness evaluation and efficiency evaluation) that a quality retrieval system should consider and possess in order to maximize user experience.

We also present several future directions in this topic that may be worthy of exploration:

- **Similarity measure**: Designing an effective similarity measure is crucial for the performance of sports play retrieval, especially learning-based similarity. As a new technique, deep learning has made a progress in many advanced tasks. It can explore the possibilities of developing a new learning-based similarity that is suitable for a concrete application in this topic.
- **Indexing technique**: Nowadays, with the proliferation of GPS devices, spatiotemporal sports data is being collected at an unprecedented speed. It is needed to develop some efficient indexing techniques to further accelerate the retrieval for large-scale databases.
- **Sub-play retrieval**: In sports area such as soccer or basketball, sports tracking systems generally record the trajectories of players and/or the ball during a long process (i.e., 90 minutes in a soccer game), and a common practice nowadays is to track the movements of players and/or the ball search for a portion/segment of play from a database of plays, with its trajectories of players and/or its trajectory of the ball similar to those and/or that of a given query play. This tasks is essentially one of considering sub-trajectory similarity [48], and it may be interesting to explore sub-play retrieval further.

REFERENCES

[1] L. Sha, P. Lucey, Y. Yue, P. Carr, C. Rohlf, and I. Matthews, "Chalkboarding: A new spatiotemporal query paradigm for sports play retrieval," in *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 2016, pp. 336–347.

[2] T. Decroos, J. Van Haaren, and J. Davis, "Automatic discovery of tactics in spatio-temporal soccer match data," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 223–232.

[3] R. Aoki, R. M. Assuncao, and P. O. Vaz de Melo, "Luck is hard to beat: The difficulty of sports prediction," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017, pp. 1367–1376.

[4] A. Bialkowski, P. Lucey, P. Carr, Y. Yue, S. Sridharan, and I. Matthews, "Large-scale analysis of soccer matches using spatiotemporal tracking data," in *Data Mining (ICDM), 2014 IEEE International Conference on*. IEEE, 2014, pp. 725–730.

[5] B.-K. Yi, H. Jagadish, and C. Faloutsos, "Efficient retrieval of similar time sequences under time warping," in *Data Engineering, 1998. Proceedings., 14th International Conference on*. IEEE, 1998, pp. 201–208.

[6] D. He, B. Ruan, B. Zheng, and X. Zhou, "Trajectory set similarity measure: an emd-based approach," in *Australasian Database Conference*. Springer, 2018, pp. 28–40.

[7] L. Sha, P. Lucey, Y. Yue, X. Wei, J. Hobbs, C. Rohlf, and S. Sridharan, "Interactive sports analytics: An intelligent interface for utilizing trajectories for interactive sports play retrieval and analytics," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 25, no. 2, p. 13, 2018.

[8] M. Di, D. Klabjan, L. Sha, and P. Lucey, "Large-scale adversarial sports play retrieval with learning to rank," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 12, no. 6, p. 69, 2018.

[9] Z. Wang, C. Long, G. Cong, and C. Ju, "Effective and efficient sports play retrieval with deep representation learning," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 499–509.

[10] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International Conference on Machine Learning*, 2014, pp. 1188–1196.

[11] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111–3119.

[12] J. Zhang, D. Wolfram, and P. Wang, "Analysis of query keywords of sports-related queries using visualization and clustering," *J. Assoc. Inf. Sci. Technol.*, vol. 60, no. 8, pp. 1550–1571, 2009.

[13] C. R., G. R., R. R., S. P., and R. S., "Ontology based automatic query expansion for semantic information retrieval in sports domain," 2012.

[14] J. Wang, C. Xu, C. E. Siong, and Q. Tian, "Sports highlight detection from keyword sequences using HMM," in *Proceedings of the 2004 IEEE International Conference on Multimedia and Expo, ICME 2004, 27-30 June 2004, Taipei, Taiwan*, 2004.

[15] Yu-Lin Kang, Joo-Hwee Lim, M. S. Kankanhalli, Chang-Sheng Xu, and Qi Tian, "Goal detection in soccer video using audio/visual keywords," in *2004 International Conference on Image Processing, 2004. ICIP '04.*, 2004.

[16] M. Xu, N. C. Maddage, C. Xu, M. S. Kankanhalli, and Q. Tian, "Creating audio keywords for event detection in soccer video," in *Proceedings of the 2003 IEEE International Conference on Multimedia and Expo, ICME 2003, 6-9 July 2003, Baltimore, MD, USA*, 2003.

[17] L. Duan, M. Xu, T. Chua, Q. Tian, and C. Xu, "A mid-level representation framework for semantic sports video analysis," in *Proceedings of the Eleventh ACM International Conference on Multimedia, Berkeley, CA, USA, November 2-8, 2003*, 2003, pp. 33–44.

[18] C. Xu, J. Wang, H. Lu, and Y. Zhang, "A novel framework for semantic annotation and personalized retrieval of sports video," *IEEE Trans. Multimedia*, 2008.

[19] H. Alt and M. Godau, "Computing the fréchet distance between two polygonal curves," *International Journal of Computational Geometry & Applications*, vol. 5, no. 01n02, pp. 75–91, 1995.

[20] L. Sha, P. Lucey, S. Zheng, T. Kim, Y. Yue, and S. Sridharan, "Fine-grained retrieval of sports plays using tree-based alignment of trajectories," *arXiv preprint arXiv:1710.02255*, 2017.

[21] M. Beetz, N. von Hoyningen-Huene, B. Kirchlechner, S. Gedikli, F. Siles, M. Durus, and M. Lames, "Aspogamo: Automated sports game analysis models," *International Journal of Computer Science in Sport*, vol. 8, no. 1, pp. 1–21, 2009.

[22] A. Bialkowski, P. Lucey, P. Carr, Y. Yue, S. Sridharan, and I. Matthews, "Identifying team style in soccer using formations learned from spatiotemporal tracking data," in *2014 IEEE international conference on data mining workshop*. IEEE, 2014, pp. 9–14.

[23] P. Lucey, A. Bialkowski, M. Monfort, P. Carr, and I. A. Matthews, ""quality vs quantity": Improved shot prediction in soccer using strategic features from spatiotemporal data," 2015.

[24] Y.-H. Chang, R. Maheswaran, J. Su, S. Kwok, T. Levy, A. Wexler, and K. Squire, "Quantifying shot quality in the nba," in *Proceedings of the 8th Annual MIT Sloan Sports Analytics Conference. MIT, Boston, MA*, 2014.

[25] A. Gionis, P. Indyk, R. Motwani *et al.*, "Similarity search in high dimensions via hashing," in *Vldb*, vol. 99, no. 6, 1999, pp. 518–529.

[26] P. Indyk and R. Motwani, "Approximate nearest neighbors: towards removing the curse of dimensionality," in *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, 1998, pp. 604–613.

[27] P. Lucey, A. Bialkowski, P. Carr, S. Morgan, I. Matthews, and Y. Sheikh, "Representing and discovering adversarial team behaviors using player roles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2706–2713.

[28] A. Miller, L. Bornn, R. Adams, and K. Goldsberry, "Factorized point process intensities: A spatial analysis of professional basketball," in *International Conference on Machine Learning*, 2014, pp. 235–243.

[29] Y. Yue, P. Lucey, P. Carr, A. Bialkowski, and I. Matthews, "Learning fine-grained spatial models for dynamic sports play prediction," in *Data Mining (ICDM), 2014 IEEE International Conference on*. IEEE, 2014, pp. 670–679.

[30] L. Chen and R. Ng, "On the marriage of lp-norms and edit distance," in *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*. VLDB Endowment, 2004, pp. 792–803.

[31] L. Chen, M. T. Özsu, and V. Oria, "Robust and fast similarity search for moving object trajectories," in *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*. ACM, 2005, pp. 491–502.

[32] S. Ranu, P. Deepak, A. D. Telang, P. Deshpande, S. Raghavan *et al.*, "Indexing and matching trajectories under inconsistent sampling rates," in *2015 IEEE 31st International Conference on Data Engineering (ICDE)*. IEEE, 2015, pp. 999–1010.

[33] X. Li, K. Zhao, G. Cong, C. S. Jensen, and W. Wei, "Deep representation learning for trajectory similarity computation," in *2018 IEEE 34th International Conference on Data Engineering (ICDE)*. IEEE, 2018, pp. 617–628.

[34] D. Yao, G. Cong, C. Zhang, and J. Bi, "Computing trajectory similarity in linear time: A generic seed-guided neural metric learning approach," in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 2019.

[35] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[36] P. Lucey, A. Bialkowski, P. Carr, Y. Yue, and I. Matthews, "How to get an open shot: Analyzing team movement in basketball using tracking data," in *Proceedings of the 8th annual MIT SLOAN sports analytics conference*. Citeseer, 2014.

[37] A. Bialkowski, P. Lucey, P. Carr, Y. Yue, and I. Matthews, "Win at home and draw away: Automatic formation analysis highlighting the differences in home and away team behaviors," in *Proceedings of 8th annual MIT sloan sports analytics conference*. Citeseer, 2014, pp. 1–7.

[38] X. Wei, L. Sha, P. Lucey, S. Morgan, and S. Sridharan, "Large-scale analysis of formations in soccer," in *2013 international conference on digital image computing: techniques and applications (DICTA)*. IEEE, 2013, pp. 1–8.

[39] M. Harmon, P. Lucey, and D. Klabjan, "Predicting shot making in basketball using convolutional neural networks learnt from adversarial multiagent trajectories," *Stat*, vol. 1050, p. 15, 2016.

[40] B. Loeffelholz, E. Bednar, and K. W. Bauer, "Predicting nba games using neural networks," *Journal of Quantitative Analysis in Sports*, vol. 5, no. 1, 2009.

[41] A. McCabe and J. Trevathan, "Artificial intelligence in sports prediction," in *Fifth International Conference on Information Technology: New Generations (itng 2008)*. IEEE, 2008, pp. 1194–1197.

[42] K. Wickramaratna, M. Chen, S.-C. Chen, and M.-L. Shyu, "Neural network based framework for goal event detection in soccer videos," in *Seventh IEEE International Symposium on Multimedia (ISM'05)*. IEEE, 2005, pp. 8–pp.

[43] 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 *International Conference on Artificial Neural Networks*. Springer, 2010, pp. 154–159.

[44] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.

[45] H. Su, K. Zheng, H. Wang, J. Huang, and X. Zhou, "Calibrating trajectory data for similarity-based analysis," in *Proceedings of the 2013 ACM SIGMOD international conference on management of data*. ACM, 2013, pp. 833–844.

[46] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*. Cambridge university press, 2008.

[47] O. Chapelle, T. Joachims, F. Radlinski, and Y. Yue, "Large-scale validation and analysis of interleaved search evaluation," *ACM Transactions on Information Systems (TOIS)*, vol. 30, no. 1, pp. 1–41, 2012.

[48] Z. Wang, C. Long, G. Cong, and Y. Liu, "Efficient and effective similar subtrajectory search with deep reinforcement learning," *arXiv preprint arXiv:2003.02542*, 2020.