Chalkboarding: A New Spatiotemporal Query Paradigm for Sports Play Retrieval

Long Sha*
QUT/Disney Research
long.sha@hdr.qut.edu.au

Patrick Lucey*
STATS LLC/Disney Research
plucey@stats.com

Yisong Yue
Caltech
yyue@caltech.edu

Peter Carr
Disney Research
peter.carr@disneyresearch.com

Charlie Rohlf
STATS LLC
crohlf@stats.com

Iain Matthews
Disney Research
iainm@disneyresearch.com

ABSTRACT

The recent explosion of sports tracking data has dramatically increased the interest in effective data processing and access of sports plays (i.e., short trajectory sequences of players and the ball). And while there exist systems that offer improved categorizations of sports plays (e.g., into relatively coarse clusters), to the best of our knowledge there does not exist any retrieval system that can effectively search for the most relevant plays given a specific input query. One significant design challenge is how best to phrase queries for multi-agent spatiotemporal trajectories such as sports plays.We have developed a novel query paradigm and retrieval system, which we call Chalkboarding, that allows the user to issue queries by drawing a play of interest (similar to how coaches draw up plays). Our system utilizes effective alignment, templating, and hashing techniques tailored to multi-agent trajectories, and achieves accurate play retrieval at interactive speeds. We showcase the efficacy of our approach in a user study, where we demonstrate orders-of-magnitude improvements in search quality compared to baseline systems.

Author Keywords

Spatiotemporal Retrieval; User Interfaces; Sports Analytics

ACM Classification Keywords

H.3.3 Information Search and Retrieval: Query Representation; H.5.2 User interfaces: Graphical User Interfaces (GUI)

INTRODUCTION

Like many data-rich domains, sports analytics is currently experiencing an explosive growth in both the quantity and granularity of data sources [35]. For instance, STATS SportVU generates location data for every player, ball, and referee at 25 Hz along with detailed logs for events such as passes, shots, fouls, etc. Given the overwhelming amount of data now being collected, the interest in effective information management and access for sports data has correspondingly grown as well.

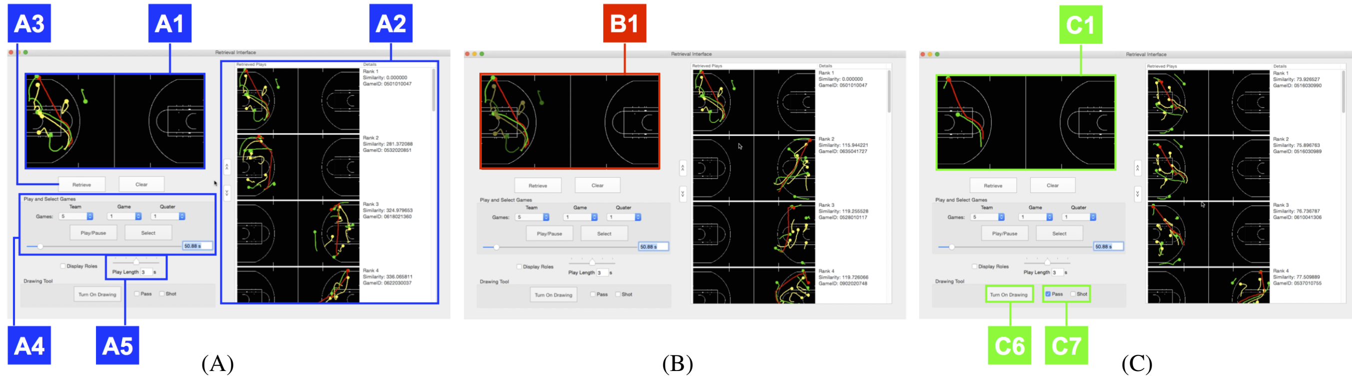
In many sports domains, an important unit of information is a "play" or a short sequence of plays. Roughly speaking, a play comprises the deliberate behavior of one or more players that

Figure 1. (a) Conventional methods for sports play access are based on "keywords" (A1), which lack specificity and requires the user to browse through a large collection before finding the specific plays of interest (A2). (b) To improve retrieval effectiveness, we instead use a "chalkboard" (B1), which is an intuitive and powerful query format that allows the user to specify plays similar to how a coach draws up plays, and can capture the rich semantics required for effective sports play retrieval (B2). The user can use the "selection tool" (B3) to select specific trajectories, or use the "drawing tool" (B4) to draw a query.

IUI 2016, March 7–10, 2016, Sonoma, CA, USA.

ACM 978-1-4503-4137-0/16/03 ...$15.00.
DOI: http://dx.doi.org/10.1145/2856767.2856772

*Part of this work was done while Sha and Lucey were at Disney Research



(A) (B) (C)

Figure 2. Our Chalkboarding query language can take three different input types. (A) Exemplar-based where a user can play an existing game in the database and choose a short segment play as query – features include: (A1) Visualization window of play containing player and ball trajectories (red=ball, green=offensive-players, yellow=defensive-players, dots=end-of-trajectory), (A2) Retrieval window which displays the most relevant plays in terms of rank, (A3) Retrieval button, (A4) Game-selection and scroll-bar to scrub through game of interest, (A5) Specify length of trajectories in seconds. (B) Manipulating exemplars where a user can select just a subset of trajectories which they deem relevant – it is the same interface except for (B1) the visualization window where the user can right-click to select the trajectories of interest. (C) Drawing plays where the user draws the play of interest like a coach would on a chalkboard – additional features include (C1) visualization window where the user draws the play of interest, (C6) draw tool selection button, (C7) toggle buttons to include actions in play (note, the user must also select the length of the play).

is designed to achieve a certain outcome, and can be thought of as short annotated trajectories of the players and the ball.

Conventional approaches to sports play access has largely focused on improving categorization of plays [38], which offers limited utility when the goal is to *retrieve specific plays*. Indeed, the lack of specificity from categorizations often results in the user to browse through a large collection of candidate plays before finding the specific play of interest. For example, in Figure 1(a), a user may search for all three point shots resulting from a pass from the right (A1).

A key challenge in sports play retrieval is how to design a query format that is both intuitive to use and rich enough to enable precise specification of an information need. Given the spatiotemporal nature of plays, it is clear that one must depart from conventional text-based query formats.

In this paper, we present a new query paradigm for sports play retrieval, which we call ***Chalkboarding***. Instead of searching plays using "keywords", we use a visual representation of player trajectories - inspired by the "x's and o's" a coach uses to draw up plays. Figure 1(b) depicts our basic setup: given an example play as the input query (B1), the system retrieves a ranked list of plays ordered by relevance to the query (B2). With a chalkboard, a user can use the "selection tool" (B3) to select trajectories for a particular example or use the "drawing tool" (B4) to draw a query. As such, Chalkboarding is an intuitive yet powerful query format that allows the user to specify plays similar to how a coach draws up plays, and can thus capture the rich semantics required for effective sports play retrieval.

We view Chalkboarding as complementary to existing categorization methods. Much like how the early work on information access for the World Wide Web was largely dichotomized into directory/taxonomy versus query/retrieval paradigms [45], we view our approach as the first instance of the query/retrieval paradigm for sports plays to complement the existing approaches for the directory/taxonomy paradigm. Note that the two types of approaches can be integrated together for improved performance (e.g., one can retrieve for only certain types of plays).

An immediate technical question that arises from our Chalkboarding query format is how to measure relevance between the query and a candidate play, i.e., how to interpret the query. For instance, sports plays are multi-agent trajectories, and so relevance estimation requires solving a potentially combinatorial alignment problem between the agents in the query and in the candidate play.

In this paper, we also present a retrieval system tailored towards retrieval of multi-agent trajectories. Our system utilizes a role representation technique based on [36, 4] that can be used to quickly and accurately find the best agent-to-agent alignments between the query and a candidate play (or between two plays) for a variety of spatiotemporal distance and similarity measures. We then leverage this representation to develop efficient templating and hashing techniques for fast indexing of a large repository of plays. Our full-stack system can thus achieve accurate retrieval at interactive speeds.

We validate our system using a variety of empirical evaluations, ranging from analyzing the accuracy/efficiency tradeoff of the various system components, to conducting a user study on the end-to-end task of sports play retrieval. Our retrieval user study demonstrates that our Chalkboarding system can achieve orders-of-magnitude improvements in search quality compared to systems based only on categorization, which showcases the efficacy of our approach.

CHALKBOARDING QUERY LANGUAGE

Our Chalkboarding query language is based on a spatiotemporal trajectories,and uses the x,y (and z if available) positions of players and the ball as the primary input. Trajectories can either be drawn by hand or extracted/selected from recorded trajectories from real games (or a mix of both). Fig-

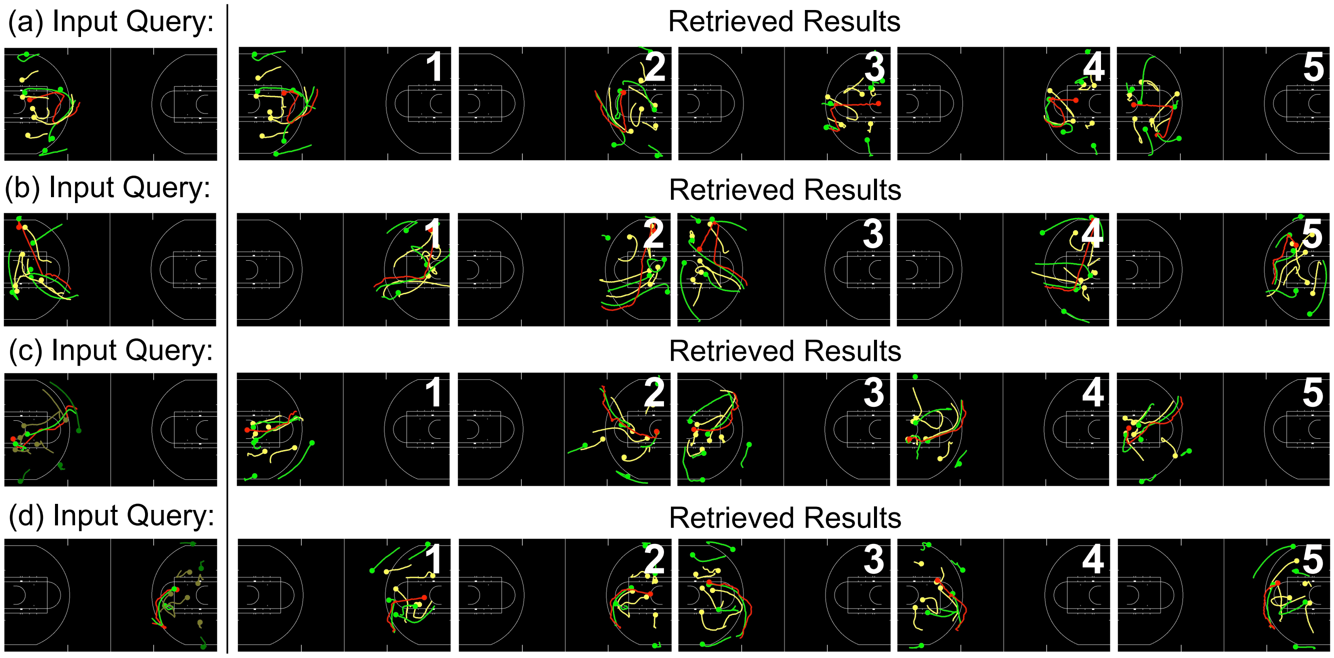


Figure 3. Different input queries using the whole game-exemplars (Query (a) and (b))and select only the trajectories of interest (Query (c) and (d)).

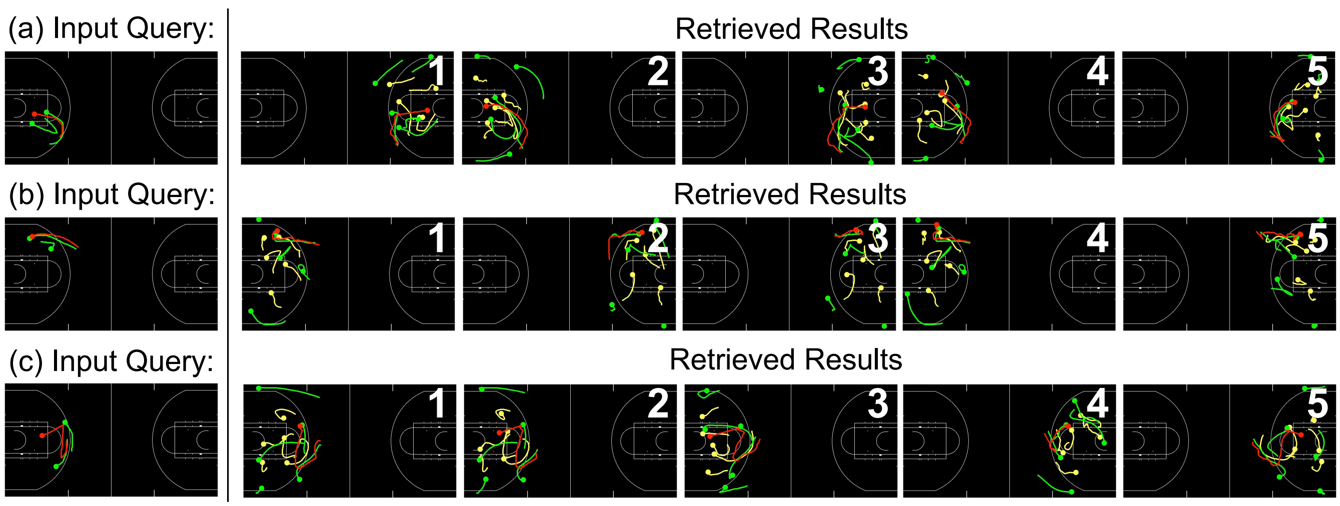


Figure 4. Different input queries using user drawn plays.

ure 2 illustrates the subtle differences between the various input queries.

Exemplar-based query: A user can use a short play sequence from an existing game as the query (Figure 2(a)). The features include: (A1) a visualization window of play containing player and ball trajectories, (A2) a retrieval window which displays the most relevant plays, (A3) a retrieval button, (A4) a game-selection and scroll-bar to scrub through game of interest until the portion of game is found, and (A5) a time-window selection. The last feature is quite important as the current system requires the user to define a time-window to the nearest second (1-5 seconds), and all searches will be constrained to look for trajectories fitting that time duration.

Manipulating exemplar-based queries: A user can also manipulate an existing play sequence before using it the query (Figure 2)(b)). The primary manipulation we consider in this paper is erasing irrelevant players (B1). More generally, one could also consider adjusting the trajectories of an existing play as well.

Drawing-based query: A user can draw a play of interest on the chalkboard. Again, the interface is similar to the other two with the only difference being that the visualization window is now a canvas in which the user draws their input play (C1). To select the drawing feature, a “draw-tool” selection button needs to be enabled (C6) and to include semantics such as “pass” or “shot” onto the trajectories, we have a toggle button for those actions included (C7).

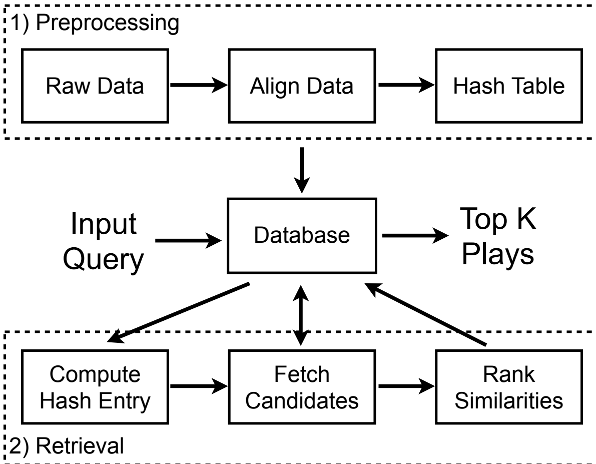


Figure 5. Our retrieval system is divided into two parts, preprocessing and retrieval. The first part aligns the raw data and builds a hash table while the second part contains a fast retrieval procedure. They are explained separately in this section.

Drawing-based query on broadcast feed: Additionally, the user can also draw on a broadcast feed of a game. To draw the play on the broadcast view is a more intuitive way of interaction. This requires an additional process of court calibration. To calibrate the court from the broadcast view with the standard overhead view, we utilize a two-step calibration method based on [12, 47] and [8]. The first step is template-based matching similar to [12, 47]. In order to do landmark matching on the court, we synthesize a relatively clean overhead view of the court by annotating a number of frames in different perspectives. Calibration is accomplished by using scale-invariant feature transform (SIFT) [34] and random sample consensus (RANSAC) [18] to match every frame to one of the templates and find the correspondence. To deal with ill-posed initial conditions, pointless calibration [8] is applied by using the calibration matrix from the previous frame.

The query types are all expressed in the format of multi-agent trajectories, and can be viewed as differing ways of constructing such trajectories. In particular, the exemplar-based and drawing-based input types represent subtractive versus additive query construction procedures: the exemplar-based queries take a full-set of multi-agent trajectories and subtract irrelevant players, whereas the drawing-based queries directly construct the trajectories of the relevant players. Note also that exemplar-based queries strictly require the user to have already identified a play of interested (e.g., from browsing existing plays), whereas drawing-based queries can be constructed without any other information.

Examples of exemplar-based and drawing-based queries are depicted in Figures 3 & 4, respectively. Note that the top-ranked retrieval plays are much more specific to the input query than those that could be retrieved from a text-based (i.e., categorization) query.

CHALKBOARDING RETRIEVAL SYSTEM

Figure 5 shows an overview of our retrieval system. To enable quick and accurate retrieval, "intelligent preprocessing" of the raw trajectory data is required. This preprocessing takes the form of using a role-based representation to enable fast alignment of trajectories, using a simple similarity measure to estimate relevance, followed by learning a hash-table to find the most relevant plays quickly.

After preprocessing, play retrieval comprises of first computing the hash entry from the query to find the most likely candidates, and then followed by local search to rank the candidate plays. Descriptions of these modules are given in the following subsections. Our full-stack system can retrieve relevant results from a repository of hundreds of thousands of plays in less than second, and can thus support interactive use cases.

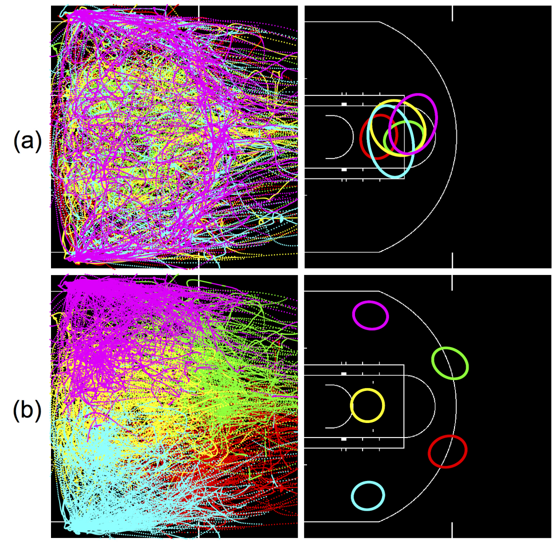


Figure 6. (a) Shows the raw trajectories (left) and the covariance (right) of each player during a quarter of play, which shows that players can be located in any part of the court. (b) However, if we align the player positions to a formation template at each frame, we show that players are spaced into a formation.

Role Representation for Fast Alignment

To measure play similarity, we employ an agent-to-agent trajectory comparison method. The advantages of using such an approach are: i) the representation is lossless (i.e., no quantization is required), ii) only a limited number of additional features are required to be stored (and not an overcomplete set of hand-crafted features which maybe hard to store in memory), iii) the representation is visual and interpretable, and iv) it allows for full interactivity with the data (i.e., users can select the precise time-window and agents of interest).

A drawback however, is the problem of permutations. An example of this problem is shown in Figure 6(a) where we show the raw trajectories and the covariances of each player across a quarter of a match. Here we can see players tend to move all around the court and not in one distinct area. In terms of matching trajectories, this means we would have to check all permutations, which in the case of basketball is $5! = 120$ (if we compare the trajectories of both teams it is 120^2).

Although exhaustively comparing all trajectories would yield the optimal agent-to-agent alignment, recent research in

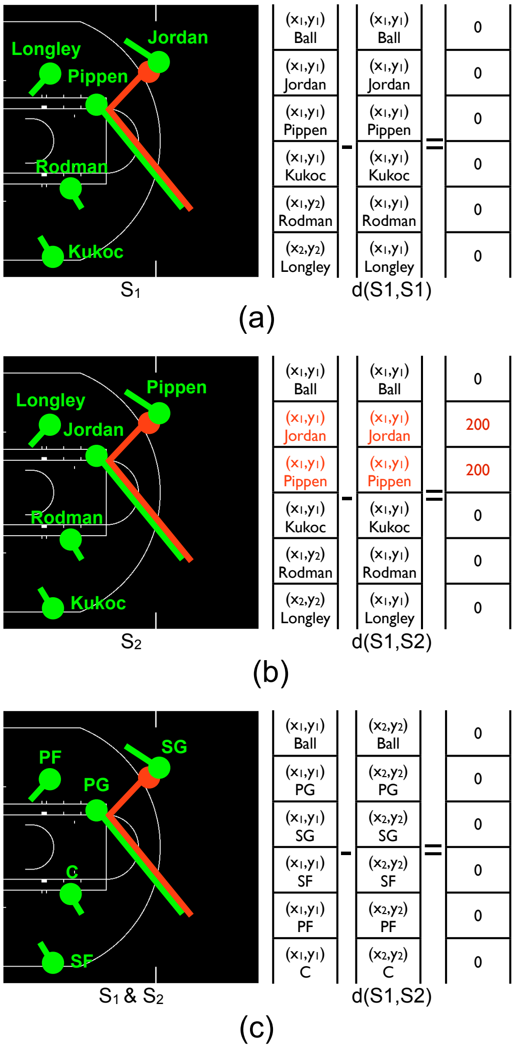


Figure 7. A toy example of the data alignment problem. (a) $\mathbf{S}_1$ is the play of interest, if we find an exactly identical play, the distance between them $d(\mathbf{S}_1, \mathbf{S}_1)$ should be zero. (b) However, if there is another same play S_2, but player Jordan and changed their roles, the distance $d(\mathbf{S}_1, \mathbf{S}_2)$ will no longer be zero due to the misalignment of these two players.

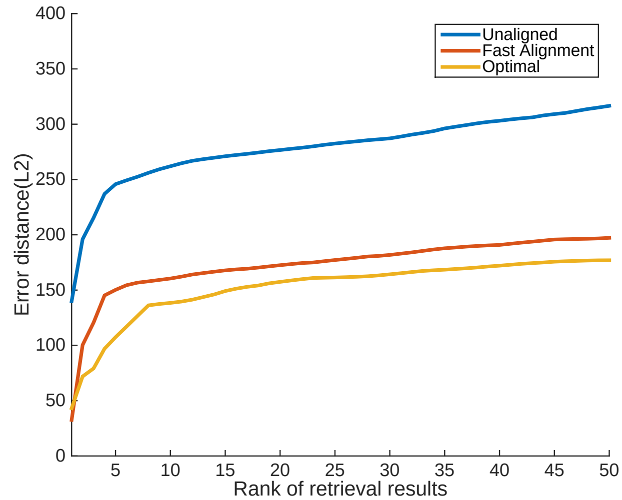


Figure 8. The plot of distance against retrieved-result ranks. Red curve is the optimal result and our role-based representation (yellow) is much closer to the red curve than identity-based representation (blue). And it is also much faster than exhaustively permuting the data.

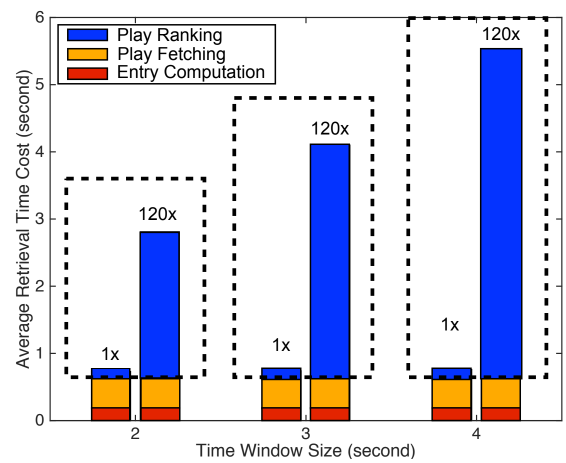


Figure 9. The time cost of our role-based representation (left bar of each pair) and exhaustive permutation (right bar of each pair). The test uses an 100,000-sample database and only the attacking team is considered.

multi-agent systems have shown that matching the trajectories to a predefined formation template can yield near-optimal agent-to-agent alignments [36, 4] in a single pass. We discuss the choice of distance metric in the next section on relevance estimation.

Figure 7 depicts a toy example that illustrates the intuition behind the formation-based approach [36, 4]. We can represent the play in Figure 7(a) as $\mathbf{S}_1 = [\mathbf{s}_{\text{ball}}, \mathbf{S}_{\text{offTeam}}]$, where each individual trajectory can be described as the vector $\mathbf{s} = [x_1, y_1, \ldots, x_F, y_F]^T$, and F is the number of frames in a play. Team behavior can then be described as the trajectories of the five players: $\mathbf{S}_{\text{offTeam}} = [\mathbf{s}_{\text{Jordan}}, \mathbf{s}_{\text{Pippen}}, \mathbf{s}_{\text{Kukoc}}, \mathbf{s}_{\text{Longley}}, \mathbf{s}_{\text{Rodman}}]$. If the team runs exactly the same play at some other point in time, then those two plays will already be pre-aligned (i.e., each player runs the exact same trajectories in both plays), and the distance between them will be zero (for any distance metric). If we instead compare $\mathbf{S}_1$ with $\mathbf{S}_2$ from Figure 7(b) using the same representation as $\mathbf{S}_1$, then the distance will not be zero as Jordan and Pippen have switched positions. However, if we discard the identity of the player and use their role information at that frame – $\mathbf{S}_{\text{offTeam}} = [\mathbf{s}_{\text{PG}}, \mathbf{s}_{\text{SG}}, \mathbf{s}_{\text{SF}}, \mathbf{s}_{\text{PF}}, \mathbf{s}_{\text{C}}]$, the distance between $\mathbf{S}_1$ and $\mathbf{S}_2$ both plays are seen as the same as the distance is zero (Figure 7(c)). In addition to normalizing for permutation, the role-representation is also robust to substitutions, and agnostic to team identity.

In terms of preprocessing, we first calculate the formation template directly from data in way similar to [4]. At each frame, we then assign each player for each team to one of five roles (PG=point-guard, SG=shooting-guard, SF=small-forward, PF=power-forward and center=C). We do this by first calculating the cost matrix between the template and the current frame snapshot. We then use the Hungarian algorithm [29] to make the assignment, which results in a single

	l_2	l_∞	DTW	Frechet	LCSS	Edit
ACC	0.773	0.755	0.787	0.741	0.430	0.412

Table 1. The average accuracy of each metric in 5-fold cross validation experiment.

role feature being added to the raw position data, which is basically just an extra column in the database.

We also run a sliding window (i.e., 1-5 seconds) and do a majority vote to determine which role that player was in for different time-windows (i.e., 1-5 seconds), which results in an additional 5 columns in the database of features, and is still very lightweight. In order to make all the plays comparable, we rotate all the plays that are conducted on the right side of court by 180 degrees so that the coordinates of all the plays are aligned.

To show that our fast alignment method approximates the optimal alignment, we randomly selected 10 queries and chose the top 50 retrieved results. Figure 8 shows the plot of the average distance against the curve of three different alignment methods (see next subsection for metric distance used). The best performing alignment method was that of the exhaustive method. The next best was our fast-alignment using the role-representation which was a close approximation to the optimal alignment. The worse performing alignment algorithm was that of identity, where the initial representation was fixed for the entire play.

Figure 9 shows the timing comparison, where we see that the role-representation alignment method is substantially faster than exhaustive alignment. This experiment was conducted on a 3.2GHz, 8GB RAM computer. It should be noted that only the attacking team is considered in this test. Had both teams (offense and defense) been involved in the retrieval task, the time cost difference would be even more significant by another order of magnitude.

Relevance Estimation

Relevance estimation is the problem of determining which plays in the data repository is most relevant to an input query, and is typically addressed by producing a ranking of decreasing similarity to the input query [42, 37]. Given that our approach compares agent-to-agent trajectories, we need a suitable similarity metric to capture the differences between plays.

We use a supervised classification experiment to evaluate which distance measure is most suitable for measuring play similarity. We asked a domain expert to annotate 1666 three-second plays into 38 classes based solely on the ball trajectory. We conducted the experiment via 5-fold cross validation and we compared six different metrics on the aligned trajectories: l_2 distance, l_∞ distance, Frechet distance, dynamic time warping (DTW), longest common subsequence (LCSS) and edit distance (Edit). For LCSS and edit distance, we use the SAX [30] to transform the time series into symbol series representation, which eliminates the effect of absolute position and focuses on the shape of trajectories. We used nearest neighbor as our classifier.

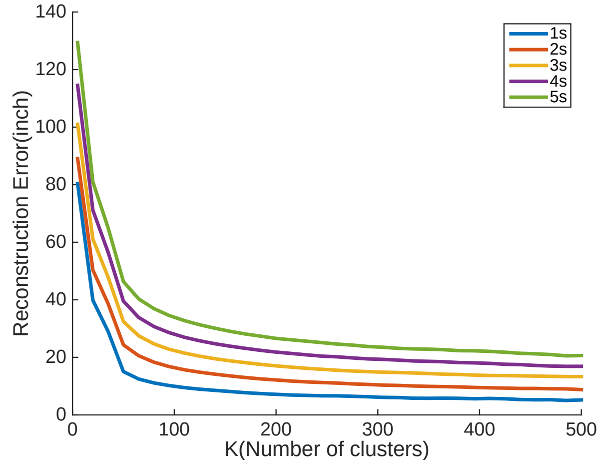


Figure 10. The reconstruction error against the number of clusters for time window sizes from 1 to 5 seconds.

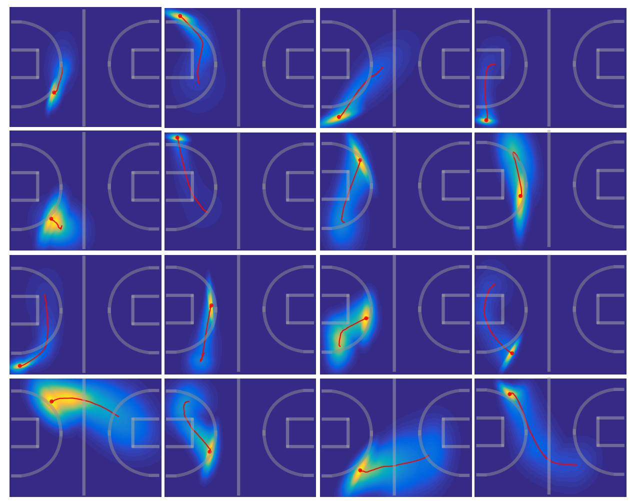


Figure 11. Some examples of the ball trajectories distribution in each cluster with 3-second time window. Red curve is the mean trajectory and the small red circles are the end points.

Table 1 shows the classification accuracy. We can that the first four metrics all worked reasonably well. To test if there were any significant differences between these four metrics we used the Mann-Whitney U significance test, which we chose as the number of examples per class were imbalanced. After running our significance tests, we found that there was no significant difference between these four. As such, we used the l_2 distance as our similarity measure as it is the easiest to deploy.

Hashing for Fast Indexing

In our dataset, we had over 600 hours worth of tracking data (∼30GB), making similarity comparisons between a query and the entire database an expensive task. For example, if we were to break the database into a series of three second plays, this would equate to 4 million different examples. If each comparison took 30ms, searching the entire dataset would take 120secs – far too long to be usable as an interactive retrieval system.

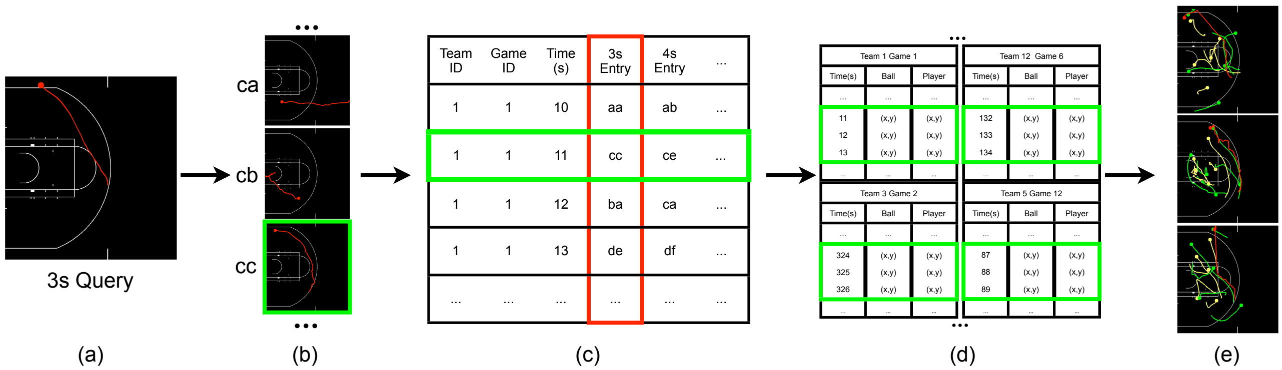


Figure 12. An example of our retrieval process: (a) 3-second query with only ball trajectory, (b) Compare to the centroid of each cluster, (c) Acquire the plays with entry cc (cluster index) from hash table, only look at 3s entry (red column) in this case, (d) According to the game index, fetch the corresponding plays from raw tracking data, (e) Find the top K nearest neighbor.

The standard approach to improve retrieval speed is to use a hash-table or some other kind of indexing [37]. Based on domain experts' opinions as well as our own observations while building the initial version of the system, we used the ball trajectory as our index feature. To learn the dictionary of indexes, we applied K-means clustering using the l_2 distance on all ball trajectories for time-windows of 1, 2, 3, 4 and 5 seconds. To choose K, we inspected the reconstruction error plots (Figure 10), which we chose to be around 60 clusters. Examples of some of the clusters for the 3 second trajectories are shown in Figure 11.

In addition to using the spatial location, ball-actions (e.g., pass, dribble, shots) were incorporated by further splitting the clusters into semantic clusters. To ensure that each cluster contained less than 1000 plays, we further divided clusters which had more than 1000 plays into sub-clusters by applying another round of K-means on the specific cluster until this was achieved (K was chosen according to the size of plays within the current cluster – e.g., if a cluster had 2000 plays, we used K=3).

In terms of preprocessing, we use each frame as the starting point of the ball trajectory and obtain the ball trajectory for various time-windows (1-5 seconds). For each time-window, we then compare that trajectory to the centroid trajectory within the hash-table for the various time-windows. We then assign the index value of the closest centroid trajectory to that frame for each time-window.

Summary of Retrieval Process

Figure 12 depicts the end-to-end process of computing similarity measures of plays in our repository against an input query. Given an input query (Figure 12(a)), we first compare its ball trajectory against the centroids of every cluster (Figure 12(b)). In this case, cluster 'cc' has the best match, and acquire the 3-second plays from that cluster using a hash table lookup (Figure 12(c) and Figure 12(d)). Finally, we perform alignment and similarity matching between the query and the plays in the cluster, and return a ranking of the most similar plays. Note that our approach can continuously run in the background and produce increasingly better rankings of retrieved plays over time, e.g., by checking the next best cluster from Figure 12(b) (although the top retrieved plays tend to be from the best cluster and so does not change).

USER STUDY EXPERIMENTS

Since Chalkboarding is designed to help users to quickly find similar plays in a huge database, we validated our system via an user study on the end-to-end task of play retrieval. The goal of the user study was to compare the search quality of rankings generated by Chalkboarding versus conventional keyword-based queries.

Baseline

The baseline system that we compare against is a keyword-based retrieval system. A basketball expert provided the keywords (action + coarse location) for each retrieval task, so that users with little prior knowledge do not need to select the keyword themselves.

Experiment Design

We selected eight retrieval tasks for our user study that are representative of basketball plays which are shown in Figure 13. For the Chalkboarding system, the first four plays are provided as exemplar-based queries which the users can use. For the remaining four queries, we only showed them the exemplar figures with selected trajectories and asked the users to draw the queries on the chalkboard as a drawing-based query.

We evaluated the retrieval quality via an interleaved evaluation, where the top thirty results from the Chalkboarding and baseline systems combined via the Team-Draft Interleaving method [9] into a single ranking (see Figure 14). The combined ranking is then embedded in the retrieval window of the Chalkboarding retrieval interface (pane B2 in Figure 1(b)), after which the user will scan the results top-down and select the relevant plays.

This setup offers two benefits: i) it is a blind paired test that can control for user- and task-specific variability, which is often more reliable and sensitive than two-sample or A/B

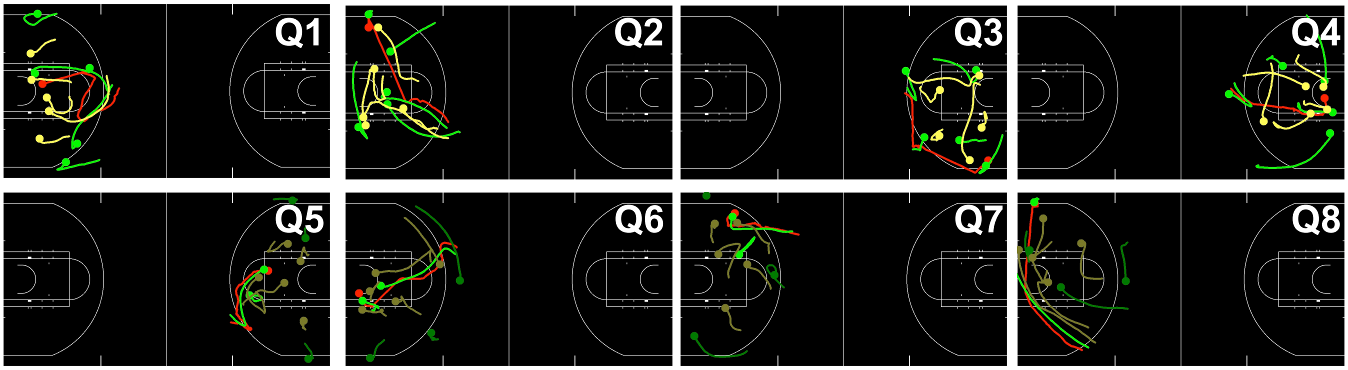


Figure 13. Depicting the eight test queries for our user study. These queries cover a wide range of plays in competitive basketball.

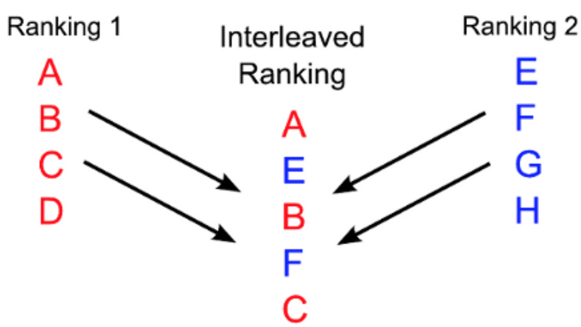


Figure 14. Depicting an interleaving of two rankings.

tests [9]; and ii) by only presenting a single ranking, interleaving allows for evaluation to be conducted in a natural usage context.

Procedure

We recruited ten volunteers with a wide range of basketball knowledge to participate in our user study. We first provided every volunteer with a ten minute introduction to ensure they understood basic basketball concepts. In particular, we used a ninth retrieval task as a demonstration for how to recognize relevant plays.

Every participant was allocated half an hour to perform all eight retrieval tasks. After the initial demonstration, all felt comfortable issuing Chalkboarding queries.

Participants used our Chalkboarding interface to retrieve each query and the result panel showed interleaved results from both systems. They were asked to scan the results top-down and highlight retrieved plays that they think are relevant to the input query. For most users, it took about ten minutes to acclimatize to the interface. After this time however, the average time it took to determine play relevance was around fifteen seconds.

Benchmark Results

Using the relevance judgments of the participants, we performed a benchmark comparison using two standard retrieval evaluation metrics: average precision [42, 37, 49] and expected reciprocal rank of the first result [10]. Let r_j denote the rank of the j-th relevant document, then the average precision of a ranking can be defined as:

$$\texttt{AvgPrec} = \frac{1}{\#rel} \sum_j Prec@r_j,$$

where $Prec@r_j$ is the precision of the top r_j items in the ranking (i.e., fraction of relevant results in the top r_j). The expected reciprocal rank is defined as:

$$\texttt{ERR} = \frac{1}{r_1},$$

which is simply the inverse of the rank of the first relevant result. Average precision is more recall-focused (i.e., places more emphasis on the rank location of all relevant results), whereas expected reciprocal rank is more sensitive to initial search time to finding the first result. For our setting, we computed both of the evaluation measures on the two virtual ranking functions embedded in the interleaved ranking, and over the pooling of both top-30 results (i.e., pooling based retrieval evaluation [37]).

	Q1	Q2	Q3	Q4	Overall
Chalkboarding	0.78	0.88	0.73	0.91	0.83
Keyword	0.06	0.17	0.24	0.09	0.14
Win / Lose	10 / 0	10 / 0	10 / 0	10 / 0	40 / 0

	Q5	Q6	Q7	Q8	Overall
Chalkboarding	0.90	0.71	0.83	0.85	0.82
Keyword	0.06	0.13	0.03	0.18	0.10
Win / Lose	10 / 0	10 / 0	10 / 0	10 / 0	40 / 0

Table 2. Comparing the mean average precision aggregated across all users for each query. Query 1-4 used exemplar-based queries in our system and 5-8 used drawing-based query. The "Win / Lose" rows show the number of users for whom Chalkboarding achieved a higher average precision. We see that Chalkboarding achieves orders-of-magnitude better retrieval quality.

Table 2 shows the results for average precision. The top two rows of the tables show the mean average precision aggregated across all ten users for each system. Recall that the first four retrieval tasks performed were using exemplar-based queries and the second four were using drawing-based

	Q1	Q2	Q3	Q4	Overall
Chalkboarding	0.83	0.85	0.63	1	0.8262
Keyword	0.03	0.06	0.14	0.04	0.07
Win / Lose	10 / 0	10 / 0	10 / 0	10 / 0	40 / 0

	Q5	Q6	Q7	Q8	Overall
Chalkboarding	0.95	0.82	0.95	0.9	0.90
Keyword	0.03	0.11	0.02	0.14	0.08
Win / Lose	10 / 0	10 / 0	10 / 0	10 / 0	40 / 0

Table 3. Comparing the expected reciprocal rank of the first relevant result across all users for each query. Comparing the mean average precision aggregated across all users for each query. Query 1-4 used exemplar-based queries in our system and 5-8 used drawing-based query. The "Win / Lose" rows show the number of users for whom Chalkboarding found a relevant result earlier in the ranking. We see that Chalkboarding achieves orders-of-magnitude better retrieval quality.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Chalkboarding	10	9	7	10	10	9	10	8
Keyword	0	1	3	0	0	1	0	2

Table 4. Showing users' reported interface preference for each query. We see that users overwhelmingly preferred Chalkboarding over the baseline system.

queries. In both settings, our Chalkboarding system substantially outperformed the baseline system. It should be noted that the difference in retrieval quality is very large for all queries.

The "Win / Lose" rows in Table 2 show the breakdown of how many individual users experienced higher average precision using Chalkboarding compared to the baseline system. We see that our Chalkboarding system wins for every user on every retrieval task.

Table 3 shows the results for expected reciprocal rank, and is structurally analogous to Table 2. We again see that our Chalkboarding system achieves significantly better performance compared to the baseline approach.

Subjective Evaluation

We also conducted two subjective evaluations. In the first evaluation, we showed participants both the keyword-based interface (i.e., Figure 1(a)) and the Chalkboard interface and asked them which interface they prefer to use for each retrieval task. All participants had previous experience with keyword-based search interfaces and naturally understood how to use our keyword-based interface. Table 4 shows the results. For most retrieval tasks, the participants unanimously preferred Chalkboarding over the baseline system. Only for queries 2, 3, 6 and 8, some users still like to use the search bar because those plays were easy to describe.

We finally asked the participants to answer a short survey regarding their user experience, and the results are shown in in Figure 15. All participants agreed that our system was more enjoyable to use because it had more functions as well as being more intuitive. In terms of effectiveness, two users with rich basketball knowledge thought the two systems were similar because they didn't think using terminology to describe

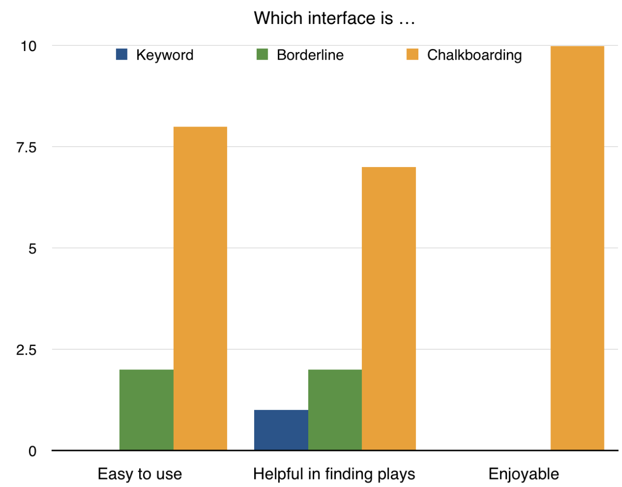


Figure 15. Showing results of our subjective survey. We see that users largely preferred the Chalkboarding system and found it easy and enjoyable to use.

a play is more complex than drawing. But for other users, instead of using domain knowledge to describe a play, drawing a play was definitely easier to find what they want. Seven participants thought Chalkboarding was more helpful for play retrieval, while the remaining three were neutral.

In our open-ended discussions about the system after the user study, the main criticism raised by the three participants who did not find Chalkboarding more helpful is that it is sometimes also important to acquire all the plays of one category. One suggestion was to have a hybrid system, where the user could use both the key-word and Chalkboard for their input query. Another criticism was that the user had to assign a role to each player, which some felt was unnecessarily complicated. We agree with this comment as this step can be automated. In future iterations, we intend to include this automated role assignment feature which will make the interface simpler.

RELATED WORK

The study of information retrieval enjoys a long and rich history in the information science and computer science communities [42, 37]. The overwhelming majority of previous work has been based on text-based or other types of tokenized query formats, which are unsuitable for sports play retrieval due to the inherent spatiotemporal nature of plays.

The most commonly studied type of retrieval and recommender systems that incorporate spatial or temporal data are those that are "location-aware", i.e., make use of information regarding the user's location to retrieve or recommend more relevant information [1, 48]. However, such approaches generally do not study new query formats.

Indeed, the use of free-form or "ad-hoc" queries in (mostly) unstructured corpora has proven to be significantly more user-friendly than more structured query types (e.g., SQL) [42, 37], and have come to dominate the commercial search industry. In a sense, our Chalkboarding approach can be viewed as a new variant of ad-hoc retrieval designed for spatiotemporal trajectory domains.

The primary complement to the query/retrieval paradigm within the broader field of information management and access is the directory/taxonomy paradigm [45]. Previous work on information access in the sports domain have largely focused on the directory/taxonomy paradigm via improving categorization of plays [38, 13, 3], which offers largely orthogonal benefits compared to the query/retrieval paradigm. However, similar to how one can construct a taxonomy of queries in web search [6], we have shown that integrating the two types of approaches can offer further benefit. Other work in sports analytics have largely focused on developing advanced metrics to evaluate player performance [19], analyzing video [32], or analyzing basic spatial patterns [39, 50].

Our approach bears affinity to other work on redesigning the interface between the human user and the data repository for various information retrieval and gathering tasks [5, 11, 16, 43]. Oftentimes, tuning system components such as the relevance estimation method results in relatively modest improvements in performance (cf. [49]), whereas designing a new interface to either accept richer inputs or produce richer outputs can lead to orders-or-magnitude improvements in overall system quality.

From the technical perspective, the primary challenge that we study is how to compare the similarity between two multi-agent trajectories. There have been ample previous work studying how to measure similarity between trajectories and time series [14, 44, 15, 21, 52], but they are largely focused on single trajectories rather than multi-agent ones. We build upon recent work that leverage a "role-based" representation [36, 46] that can compactly and efficiently characterize group behavior and formation in the sports domains.

The role representation is used primarily for alignment purposes and leaves open the question of what similarity measure to use for comparing individual pairs of aligned trajectories. One popular line of research focused on elastic measure that address warping and shifting effects in space and/or time [14, 2, 27, 28, 41, 31], while another line of research focused on finding the most simlar or dissimilar points in two trajectories in order to ensure some notion of robustness (e.g., via the Hausdorff distance) [33, 25]. As the latter is designed to measure distance between polygons and does not take the direction of trajectories into consideration [52], in this paper, we test some representative algorithms from the former category to find an ideal one for our task.

All modern retrieval systems require fast indexing in order to quickly search through a large repository, and one popular approach (which we also adopt) is to use a hash table [23, 51]. In different domains, the hash function is designed differently for specific applications. In general, the goal of hashing is to minimize time cost. Our approach is built on the concept of locality sensitive hashing (LSH) [24, 20], which is designed to place similar samples into a same address. Such an approach has also been applied in other settings where a similarity measure or ranking is required [26, 40].

DISCUSSION OF LIMITATIONS & FUTURE WORK

In a sense, our approach can be thought of as the simplest approach that can support effective sports play retrieval, and motivates many directions for future work. For instance, we can incorporate more sophisticated categorization techniques in order to further refine our query language (i.e., further combine the query/retrieval and directory/taxonomy paradigms). From an interface design standpoint, one obvious area for improvement is spatial manipulation of existing or pre-drawn trajectories (i.e., slightly translating or rotating an existing trajectory).

There are other important information access tasks beyond retrieval, most notably including summarization and personalized curation [11, 43]. A typical interface for the summarization setting is to construct clusters of results and describe the relationships between clusters. Chalkboarding can also be used to improve categorization systems by creating exemplars for clustering purposes.

From a system design standpoint, it would be highly beneficial to further improve retrieval speeds. For instance, many applications designed on top of web search require issuing multiple search queries, which is impractical given our current system. It is possible that using a hierarchical hashing/indexing technique and/or coresets [22, 17] can substantially improve retrieval speed.

From a relevance estimation standpoint, our choice of distance measures can be improved. For instance, one of the early breakthroughs in conventional information retrieval is the concept of "inverse document frequency" [42], which essentially states that tokens that appear in many documents in the corpuse (e.g., stopwords such as "the") are not as indicative of relevance more rare words. It would be interesting to incorporate this concept into our our trajectory distance measures. More generally, one can consider applying machine learning to develop even more accurate measures given appropriate training data [49, 7].

Another interesting direction is to incorporate more spatial regularities into the query processing. For example, understanding spatial equivalence classes [39, 50] can enable the retrieval system to understand that the left corner three point shot is (almost) equivalent to the right corner three point shot, which can further improve the accuracy of the system.

Beyond sports domains, query formats such as Chalkboarding could be applied to a wide range of spatiotemporal retrieval settings. Perhaps the most natural data domain is other types of behavioral tracking data, such as animal behavior.

CONCLUSION

We have presented Chalkboarding, which is a new query format that can naturally capture complex semantics of multi-agent trajectories for sports play retrieval. We have also presented a retrieval system tailor towards efficient and accurate sports play retrieval. Our full-stack system can achieve accurate retrieval at interactive speeds, and we demonstrate its effectiveness in a retrieval user study where our approach achieves orders-of-magnitude improvement in retrieval effectiveness over baseline methods.

REFERENCES

1. Bennett, P. N., Radlinski, F., White, R. W., and Yilmaz, E. Inferring and using location metadata to personalize web search. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, ACM (2011), 135–144.
2. Berndt, D. J., and Clifford, J. Using dynamic time warping to find patterns in time series. In *KDD workshop*, Seattle, WA (1994), 359–370.
3. Bialkowski, A., Lucey, P., Carr, P., Yue, Y., and Matthews, I. "win at home and draw away": Automatic formation analysis highlighting the differences in home and away team behaviors. In *MIT Sloan Sports Analytics Conference (SSAC)* (2014).
4. Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., and Matthews, I. Large-scale analysis of soccer matches using spatiotemporal tracking data. In *ICDM* (2014).
5. Brandt, C., Joachims, T., Yue, Y., and Bank, J. Dynamic ranked retrieval. In *ACM International Conference on Web Search and Data Mining (WSDM)*, ACM (2011), 247–256.
6. Broder, A. A taxonomy of web search. *ACM Sigir forum 36*, 2 (2002), 3–10.
7. Burges, C. J. From ranknet to lambdarank to lambdamart: An overview. Tech. Rep. MSR-TR-2010-82, Microsoft Research, 2010.
8. Carr, P., Sheikh, Y., and Matthews, I. Point-less calibration: Camera parameters from gradient-based alignment to edge images. In *IEEE Workshop on Applications of Computer Vision (WACV)*, IEEE (2012), 377–384.
9. Chapelle, O., Joachims, T., Radlinski, F., and Yue, Y. Large-scale validation and analysis of interleaved search evaluation. *ACM Transactions on Information Systems (TOIS) 30*, 1 (2012), 6:1–6:41.
10. Chapelle, O., Metlzer, D., Zhang, Y., and Grinspan, P. Expected reciprocal rank for graded relevance. In *ACM Conference on Information and Knowledge Management (CIKM)*, ACM (2009), 621–630.
11. Chau, D. H., Kittur, A., Hong, J. I., and Faloutsos, C. Apolo: making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2011), 167–176.
12. Chen, J., and Carr, P. Mimicking human camera operators. In *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, IEEE (2015), 215–222.
13. Chen, S., Feng, Z., Lu, Q., Mahasseni, B., Fiez, T., Fern, A., and Todorovic, S. Play type recognition in real-world football video. In *IEEE Winter Conference on Applications of Computer Vision (WACV)* (2014).
14. Chen, Y., Nascimento, M., Ooi, B. C., and Tung, A. Spade: On shape-based pattern detection in streaming time series. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on* (April 2007), 786–795.
15. Eichmann, P., and Zgraggen, E. Evaluating subjective accuracy in time series pattern-matching using human-annotated rankings. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, ACM (2015), 28–37.
16. El-Arini, K., and Guestrin, C. Beyond keyword search: discovering relevant scientific literature. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, ACM (2011), 439–447.
17. Feldman, D., Faulkner, M., and Krause, A. Scalable training of mixture models via coresets. In *Neural Information Processing Systems (NIPS)* (2011), 2142–2150.
18. Fischler, M. A., and Bolles, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM 24*, 6 (1981), 381–395.
19. Franks, A., Miller, A., Bornn, L., and Goldsberry, K. Counterpoints: Advanced defensive metrics for nba basketball. In *MIT Sloan Sports Analytics Conference (SSAC)* (2015).
20. Gionis, A., Indyk, P., and Motwani, R. Similarity search in high dimensions via hashing. In *Conference on Very Large Databases (VLDB)*, vol. 99 (1999), 518–529.
21. Gudmundsson, J., Laube, P., and Wolle, T. Computational movement analysis. In *Springer handbook of geographic information*. Springer, 2012, 423–438.
22. Har-Peled, S., and Mazumdar, S. On coresets for k-means and k-median clustering. In *ACM Symposium on Theory of Computing (STOC)*, ACM (2004), 291–300.
23. Huston, S., Culpepper, J. S., and Croft, W. B. Indexing word sequences for ranked retrieval. *ACM Transactions on Information Systems (TOIS) 32*, 1 (2014), 3.
24. Indyk, P., and Motwani, R. Approximate nearest neighbors: towards removing the curse of dimensionality. In *ACM Symposium on Theory of Computing (STOC)*, ACM (1998), 604–613.
25. Junejo, I. N., Javed, O., and Shah, M. Multi feature path modeling for video surveillance. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, IEEE (2004), 716–719.
26. Kafai, M., Eshghi, K., and Bhanu, B. Discrete cosine transform locality-sensitive hashes for face retrieval. *Multimedia, IEEE Transactions on 16*, 4 (2014), 1090–1103.

27. Keogh, E., and Ratanamahatana, C. A. Exact indexing of dynamic time warping. *Knowledge and information systems 7*, 3 (2005), 358–386.

28. Keogh, E. J., and Pazzani, M. J. Scaling up dynamic time warping for datamining applications. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM (2000), 285–289.

29. Kuhn, H. W. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly 2*, 1-2 (1955).

30. Lin, J., Keogh, E., Lonardi, S., and Chiu, B. A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, ACM (2003), 2–11.

31. Listgarten, J., Neal, R. M., Roweis, S. T., and Emili, A. Multiple alignment of continuous time series. In *Neural Information Processing Systems (NIPS)* (2004), 817–824.

32. Liu, T.-Y., Ma, W.-Y., and Zhang, H.-J. Effective feature extraction for play detection in american football video. In *IEEE Multimedia Modelling Conference (MMM)*, IEEE (2005), 164–171.

33. Lou, J., Liu, Q., Tan, T., and Hu, W. Semantic interpretation of object activities in a surveillance system. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 3, IEEE (2002), 777–780.

34. Lowe, D. G. Object recognition from local scale-invariant features. In *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, Ieee (1999), 1150–1157.

35. Lowe, Z. Lights, cameras, revolution. Grantland, March 2013.

36. Lucey, P., Bialkowski, A., Carr, P., Morgan, S., Matthews, I., and Sheikh, Y. Representing and discovering adversarial team behaviors using player roles. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2013).

37. Manning, C. D., Raghavan, P., Schütze, H., et al. *Introduction to information retrieval*, vol. 1. Cambridge university press Cambridge, 2008.

38. McQueen, A., Wiens, J., and Guttag, J. Automatically recognizing on-ball screens. In *MIT Sloan Sports Analytics Conference (SSAC)* (2014).

39. Miller, A., Bornn, L., Adams, R., and Goldsberry, K. Factorized point process intensities: A spatial analysis of professional basketball. In *International Conference on Machine Learning (ICML)* (2014).

40. Qin, J., Liu, L., Yu, M., Wang, Y., and Shao, L. Fast action retrieval from videos via feature disaggregation. In *BMVC*, BMVC (2015).

41. Sakurai, Y., Yoshikawa, M., and Faloutsos, C. Ftw: fast similarity search under the time warping distance. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, ACM (2005), 326–337.

42. Salton, G., and McGill, M. J. *Introduction to modern information retrieval*. McGraw-Hill, Inc., 1986.

43. Shahaf, D., Guestrin, C., and Horvitz, E. Metro maps of science. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, ACM (2012), 1122–1130.

44. Toohey, K., and Duckham, M. Trajectory similarity measures. *SIGSPATIAL Special 7*, 1 (2015), 43–50.

45. Wall, A. History of search engines: From 1945 to google today.

46. Wei, X., Sha, L., Lucey, P., Morgan, S., and Sridharan, S. Large-scale analysis of formations in soccer. In *Digital Image Computing: Techniques and Applications (DICTA), 2013 International Conference on*, IEEE (2013), 1–8.

47. Wen, P., Cheng, W., Wang, Y., Chu, H., Tang, N., and Liao, H. Court reconstruction for camera calibration in broadcast basketball videos. *Visualization and Computer Graphics, IEEE Transactions on PP*, 99 (2015), 1–1.

48. Yahi, A., Chassang, A., Raynaud, L., Duthil, H., and Chau, D. H. P. Aurigo: an interactive tour planner for personalized itineraries. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, ACM (2015), 275–285.

49. Yue, Y., Finley, T., Radlinski, F., and Joachims, T. A support vector method for optimizing average precision. In *ACM Conference on Information Retrieval (SIGIR)*, ACM (2007), 271–278.

50. Yue, Y., Lucey, P., Carr, P., Bialkowski, A., and Matthews, I. Learning fine grained spatial models for dynamic sports play prediction. In *IEEE International Conference on Data Mining (ICDM)* (2014).

51. Zhang, S., Yang, M., Cour, T., Yu, K., and Metaxas, D. N. Query specific rank fusion for image retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 37*, 4 (2015), 803–815.

52. Zhang, Z., Huang, K., and Tan, T. Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3, IEEE (2006), 1135–1138.