Proceedings of the IASTED International Conference
Artificial Intelligence and Applications (AIA 2014)
February 17 - 19, 2014 Innsbruck, Austria

QUALITATIVE SPATIO-TEMPORAL REPRESENTATION FOR EVENT EXTRACTION FROM VIDEO DATA OF FOOTBALL GAMES

Masaki Sakaida
School of Science & Technology
Kwansei Gakuin University
2-1, Gakuen, Sanda, JAPAN
email: bfd76635@kwansei.ac.jp

Takanori Kiyose
School of Science & Technology
Kwansei Gakuin University
2-1, Gakuen, Sanda, JAPAN
email: takanori0572@kwansei.ac.jp

Kazuko Takahashi
School of Science & Technology
Kwansei Gakuin University
2-1, Gakuen, Sanda, JAPAN
email: ktaka@kwansei.ac.jp

ABSTRACT
We describe a qualitative representation of the spatial relations between extracted regions of video data, and discuss event occurrence based on this representation. We use video footage of football games and investigate a formalization to determine whether an event has occurred, specifically a pass or shot for a goal. We represent mereological and directional relations of regions in each frame based on extracted regions of objects, and determine event occurrence from the sequence of these relations. This qualitative spatio-temporal method reduces computational complexity and provides clear semantics defining an event.

KEY WORDS
Qualitative spatial reasoning, spatio-temporal representation, event extraction, knowledge representation

1 Introduction

Recent advances in computer performance provide scope to manage large amounts of dynamic image data. With the growth in the amount of video data that has accompanied the rise of websites such as YouTube, there has been growing interest in the development of efficient search methods that allow users to find video data of interest and the development of systems that can describe the contents of video images. Regarding sports videos, it may be desirable to automatically detect highlights or automatically generate an outline of events. However, the amount of data in such videos is so large that it is not efficient to directly search the dataset. In this case, it may be beneficial to give annotations that describe objects or events that occur in the video, and use these metadata to search the footage. Such tagging is typically performed manually, and events are extracted from this annotated sequence. Instead of tagging, there are a number of methods for event extraction directly from spatial data by tracing the positions of moving objects. In these studies, the Hidden Markov Model (HMM) is frequently constructed from temporal changes in the positions of objects, and probabilistic statistical approaches are used, where numerical data is used to represent the positions of objects (e.g., [1]).

In this work, we take a different approach. We provide a qualitative representation of positional relations between objects and determine the occurrence of events on sequences of these representations. The qualitative treatment reduces computational complexity and gives clear semantics, since it uses symbolic data. It is advantageous in these points compared to approaches that use numerical data.

Qualitative reasoning or qualitative physics is a method which has long been studied in artificial intelligence (AI) [2]. It is used to characterize qualitative physical phenomena, such as the movement of objects. In qualitative reasoning, precise values of numerical data are not used; instead, qualitative data that indicate a change in aspects are used.

Qualitative reasoning on spatial data is called Qualitative Spatial Reasoning (QSR). It is a method that treats figures or images qualitatively, by extracting the information relevant to a user such as position, size, direction and so on [3, 4, 5].

A system that incorporates spatial relationships with dynamics is called a qualitative spatio-temporal reasoning (QSTR) system. Several frameworks for QSTR have been proposed [5] and methods for using QSTR frameworks for event extraction from video data have been studied [6, 7].

In these previous studies, either isolated moving objects or the camera position were assumed to be stable, even if there were multiple objects interacting with each other. Furthermore, objects have generally been extracted as rectangles. However, video datasets often include multiple moving objects, and the viewpoint of the camera is also moving. In addition, it may not be suitable to use rectangles to represent regions or objects, depending on the viewpoint. The methods reported thus far are not sufficient to handle such video data.

In this paper, we describe the extraction of video data of football games using a qualitative representation to characterize events. Specifically, for a given pair of extracted regions of objects in the 2D image data for each frame, we represent their relative positions based on two aspects: mereological relations and directional relations. The former refers to whether two objects occupy a common portion in a 2D plane, and the latter refers to which direction a given object is located in with respect to the other one.

Consider video footage of a football player and a ball. We represent the positional relations as follows: the ball is

DOI: 10.2316/P.2014.816-013

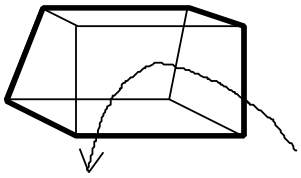


Figure 1. Trajectory of a ball on shooting

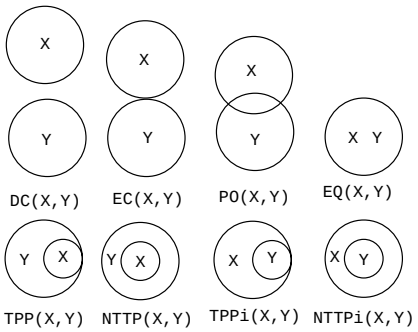


Figure 2. Fundamental relations of RCC-8

to the left of the player, and is separated so that there is no contact and no overlap between the images to the two objects. In a subsequent frame, there is some overlap of the two objects, and then the ball moves to the right of the player and the two objects are separated. In this case, we cannot determine whether the player made contact with the ball only from such a sequence of qualitative representations.

Moreover, we can determine that a player took a shot if the ball moves towards the goal area in three-dimensional (3D) space. This is more difficult using a projection onto the 2D plane. Consider the region of the goal in a 2D projection. When a player scores a goal, the region of a ball and that of a goal have a common spatial part. However, this is not sufficient in 2D, since the ball may move in the foreground or background of the goal area. For example, consider the 2D representation of the trajectory shown in Figure 1. From this, we cannot determine whether the ball entered the region of the goal. Therefore, we must define the success of a goal using the conditions that the ball passes through the ground of the goal area or a player handles it in the goal area after the ball enters the goal region. To achieve this, we extract not only the goal region but the ground part of the goal region, and represent the relation of the ground region with that of a player and the ball. In this case, we have to handle not only shapes of rectangles but polygons.

In this paper, we describe a qualitative representation of the spatial relations between the extracted regions and determine event occurrences. We evaluate our method to actual video data. Our final goal is to construct a system that automatically extracts objects from video data, generates a qualitative representation of their relations, and determines the occurrence of events from a sequence of these representations.

To identify moving objects, we use techniques commonly used in image-processing tools and pattern-recognition methods (e.g., [8]); we do not describe this in detail as it is outside the scope of this paper.

This paper is organized as follows. In Section 2, we describe the theories on which our representation is based. In Section 3, we present the spatio-temporal representation of the relations between objects and the definition of an event. In Section 4, we evaluated our method using video data and discuss the limitations of our method. In Section 5, we compare it with the related works. In Section 6, we show conclusions and future works.

2 Basic Theories

We describe the theories Region Connection Calculus (RCC) [9] and Interval Algebra (IA) [10], which our formalization is based on. We represent the relationships between objects based on RCC and consider the time interval in which an event occurs with these relationships based on IA.

Region Connection Calculus (RCC) is the theory of QSR which abstractly describes regions by their possible relations to each other. Regions are non-empty regular, closed subsets of a topological space, and can consist of more than one piece. Several versions of RCC exist depending on the granularity of classification of relations. RCC-8, for example, has eight fundamental relations is one of them. Figure 2 shows the relations of RCC-8: disconnected (DC), externally connected (EC), equal (EQ), partially overlapping (PO), tangential proper part (TPP), tangential proper part inverse (TPPi), non-tangential proper part (NTPP), and non-tangential proper part inverse (NTPPi). These relations are pairwise disjoint and jointly exhaustive.

Interval Algebra (IA) is a calculus for temporal reasoning. Relations between intervals are formalized as sets of basic relations: $before$, $meets$, $overlaps$, $starts$, $during$,$finishes$ and $equals$ (Figure 3). For a pair of time intervals I_1 and I_2, exactly one of these relations holds. The expressions $begin(I)$ and $end(I)$ denote the beginning and end of a time interval I. The symbol '$=$' denotes that two events occur at the same time. Let I_1 and I_2 be time intervals where $meets(I_1, I_2)$. We can create an interval $I = I_1 + I_2$ where $begin(I) = begin(I_1), end(I) = end(I_2)$.

3 Qualitative Representation

We describe the spatio-temporal representation of the relations between objects and the definition of an event. The ball and players are dynamic objects, whereas the goals are static objects. We describe the positional relations between

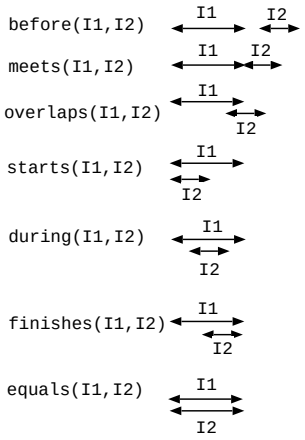


Figure 3. Relations of two intervals

TL	T	TR
L	M	R
BL	B	BR

Figure 4. Nine areas

each player and the ball, the goal and the ball, and the goal and each player.

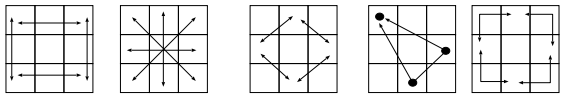
3.1 Universe division

Let $universe$ be a closed region corresponding to the entire image data in the video at an instant. The regions of the ball and the players are taken to be upright rectangles. For each player, we divide the universe into nine areas, regarding the player as a main center M. The surrounding eight areas are T, TR, R, BR, B, BL, L, and TL, clockwise from the top (Figure 4). We represent the relative position of each player and the ball using the areas that the ball occupies.

Let $S_P = \{T, TR, R, BR, B, BL, L, TL\}$. We define four subsets of S_P so that the areas in the same direction are included in that subset: $Tp = \{TR, T, TL\}$, $Rt = \{TR, R, BR\}$, $Lt = \{TL, T, BL\}$ and $Bt = \{BR, B, BL\}$. We call these subsets the *classes of direction*.

We define three binary relations and two ternary relations over S_P as follows: these are used to determine whether a player has made contact with the ball.

- For a pair of $a, a' \in S_P$, if they are included in the same class of direction, then a and a' are *in the same direction* and are denoted by $same(a, a')$ (Figure 5(a)).
- $Opposite = \{(T, B), (R, L), (TR, BL), (TL, BR)\}$. For a pair of $a, a' \in S_P$, if $(a, a') \in Opposite$, then they are *in the opposite direction* and are denoted by $opposite(a, a')$ (Figure 5(b)).
- $Diagonal = \{(T, R), (R, B), (B, L), (L, T)\}$. For a pair of $a, a' \in S_P$, if $(a, a') \in Diagonal$, then they are *in the diagonal direction* and denoted by $diagonal(a, a')$ (Figure 5(c)).
- $Knight = \{(TL, R, B), (TR, B, L), (BR, L, T), (BL, R, T)\}$. For areas $a, a', a'' \in S_P$, if $(a, a', a'') \in Knight$, then they are *in the knight direction* and denoted by $knight(a, a', a'')$ (Figure 5(d)).
- $Corner = \{(TL, L, T), (TR, T, R), (BR, R, B), (BL, B, L)\}$. For areas $a, a', a'' \in S_P$, if $(a, a', a'') \in Corner$, then they are *in the corner direction* and denoted by $corner(a, a', a'')$ (Figure 5(e)).



(a)same (b)opposite (c)diagonal (d)knight (e)corner

Figure 5. Relations over areas

For any $a, a', a'' \in S_P$ the following properties hold.
(P1) $same(a, a') \rightarrow same(a', a)$
(P2) $same(a, a') \wedge same(a', a'') \rightarrow same(a, a'')$
(P3) $opposite(a, a') \rightarrow opposite(a', a)$
(P4) $diagonal(a, a') \rightarrow diagonal(a', a)$
(P5) $diagonal(a, a') \wedge diagonal(a', a'') \rightarrow opposite(a, a'')$
(P6) $knight(a, a', a'') \rightarrow knight(a, a'', a')$
(P7) $knight(a, a', a'') \rightarrow diagonal(a', a'')$
(P8) $corner(a, a', a'') \rightarrow corner(a, a'', a')$
(P9) $corner(a, a', a'') \rightarrow diagonal(a', a'')$

We also define a relationship between sets of areas. Let A, A' and A'' be subsets of $S_P \cup \{M\}$.

- If $\forall a, a' \in A \cup A'$ s.t. $same(a, a')$, then A and A' are *in the same direction* and denoted by $same(A, A')$.
- If $\exists a \in A, \exists a' \in A'$ s.t. $opposite(a, a')$, then A and A' are *in the opposite direction* and denoted by $opposite(A, A')$.
- If $\exists a \in A, \exists a' \in A'$ s.t. $diagonal(a, a')$, then A and A' are *in the diagonal direction* and denoted by $diagonal(A, A')$.
- If $\exists a \in A$, $\exists a' \in A'$, $\exists a'' \in A''$ s.t. $knight(a, a', a'')$, then (A, A', A'') is *in the knight direction* and denoted by $knight(A, A', A'')$.
- If $\exists a \in A$, $\exists a' \in A'$, $\exists a'' \in A''$ s.t. $corner(a, a', a'')$, then (A, A', A'') is *in the corner direction* and denoted by $corner(A, A', A'')$.

3.2 Relation between player and ball

Let $Players$ be a set of players. A player is represented by a pair of a team and his/her ID, $(Team, ID)$.

Mereological relations between a player and the ball are represented using three predicates: DC (disconnected), EC (externally connected), and O (overlapped). We use C (connected) if either DC or EC holds. In contrast to RCC-8, we do not have to use EQ, $TPPi$, or $NTPPi$ relations, since the ball is much smaller than the players. Let $Areas \subseteq S_P \cup \{M\}$ be a set of areas in which the inner part of the ball exists.

- If the region corresponding to the ball does not have a common part with the area M with respect to a player P, then we denote $DC(P, Areas)$.
- If the region corresponding to the ball is touched either by a line or a point to the area M with respect to a player P, then we denote $EC(P, Areas)$.
- If the region corresponding to the ball has a common part with the area M with respect to a player P, then we denote $O(P, Areas)$.

Note that if $DC(P, Areas)$ or $EC(P, Areas)$ holds, then $|Areas| \leq 2$; if $O(P, Areas)$ holds, then $|Areas| = 1$ or 3.

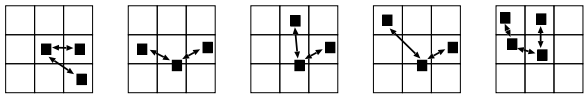
For the ball and each player, we represent the mereological relations using these predicates, and represent directional relations using the above notion of areas.

We represent a state in each instant (i.e., frame of video) by $r(P, Areas, T)$ where $r \in \{DC, EC, O\}$, $P \in Players$, $Areas \subseteq S_A \cup \{M\}$ is a set of areas with respect to P where the inner part of the ball exists, and T is an instant at which $r(P, Areas)$ holds. For any P and T, three relations DC, EC, and O are jointly exhaustive and pairwise disjoint. that is, $DC(P, A_1, T) \vee EC(P, A_2, T) \vee O(P, A_3, T)$, $\neg(DC(P, A_1, T) \wedge EC(P, A_2, T))$, $\neg(EC(P, A_1, T) \wedge O(P, A_2, T))$ and $\neg(O(P, A_1, T) \wedge DC(P, A_2, T))$ hold.

Let $T_0, T_1, \ldots, T_{n+1}$ be a successive time sequence. Consider a sequence of relations on a player P, $r_0(P, A_0, T_0), \ldots, r_{n+1}(P, A_{n+1}, T_{n+1})$. If both $r_i = r_{i+1}$ and $A_i = A_{i+1}$ hold for any i $(1 \leq i \leq n-1)$, but not hold for $i = 0, n$, then we take $T_1, \ldots, T_n$ as one interval I. For an interval I, $r(P, A, I)$ indicates that $r(P, A)$ holds during I.

3.3 Contact with the ball

If P makes contact with the ball in time interval I, we denote this by $contact(P, I)$. When $DC(P, A, I)$ holds, $\neg contact(P, I)$ obviously holds; when $C(P, A, I)$ holds, we can determine whether or not $contact(P, I)$ holds in most cases by considering the change in the direction of the motion of the ball as follows.



(a) (b) (c) (d) (e)

Figure 6. Contact conditions

[Event of contact]

Consider the following sequence of spatial relations of player and ball: $DC(P, A_1, I_1)$, $C(P, A_2, I_2)$, ..., $C(P, A_{n-1}, I_{n-1})$, $DC(P, A_n, I_n)$, where $meets(I_i, I_{i+1})$ $(1 \leq i \leq n-1)$ holds. We call this sequence *P's sequence on contact*, and a sum of intervals $I_2 + \ldots + I_{n-1}$, *P's holding time*. We have the following conditions for determining the occurrences of the event of contact for a sequence of a player's contact, that are formalized as follows.

Let I be a P's holding time.

- If $same(A_1, A_n)$, then $contact(P, I)$. (Figure 6(a)).
- If A_1 is a singleton, and there exists $i (1 < i < n)$ such that $O(P, A_i, I_i)$ is satisfied and one of the following conditions holds, then $contact(P, I)$ holds.
  - $opposite(A_1, A_n) \wedge diagonal(A_1, A_i)$ (Figure 6(b))
  - $(diagonal(A_1, A_i) \wedge opposite(A_i, A_n)) \vee (diagonal(A_n, A_i) \wedge opposite(A_i, A_1))$ (Figure 6(c))
  - $knight(A_n, A_i, A_1) \vee knight(A_1, A_i, A_n)$ (Figure 6(d))
- If A_1 is a singleton, $(\ DC(P, A_{n+1}, I_{n+1}) \wedge meets(I_n, I_{n+1}) \wedge corner(A_{n+1}, A_1, A_n)\ ) \vee (\ DC(P, A_0, I_0) \wedge meets(I_0, I_1) \wedge corner(A_1, A_0, A_n)\ )$, then $contact(P, I)$ holds (Figure 6(e)).
- If any of the above conditions is not satisfied and $opposite(A_1, A_n)$ holds, then $\neg contact(P, I)$ holds.
- Otherwise, we cannot determine whether or not $contact(P, I)$ holds.

[Event of transfer]

Let I_1 be an interval in which $contact(P_1, I_1)$ holds. If conditions (i) and (ii) hold, then the event of $transfer$ from P_1 to P_2 occurs during $I_1 + I_2 + I_3$.

(i) $meets(I_1, I_2) \wedge meets(I_2, I_3)$ and $|begin(I_3) - end(I_1)|$ is the minimum for all $P_2 \in Players$.

(ii) $contact(P_1, I_1) \wedge \neg contact(P_2, I_1) \wedge \neg contact(P_1, I_2) \wedge \neg contact(P_2, I_2) \wedge \neg contact(P_1, I_3) \wedge \neg contact(P_2, I_3)$

[Event of pass]

If the event that $transfer$ from P_1 to P_2 occurs during interval I, and $P_1 = (Team_1, ID_1) \wedge P_2 =$

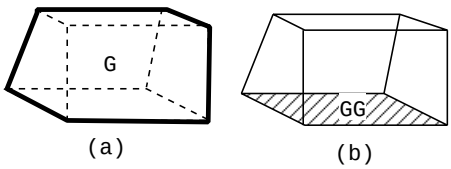


Figure 7. The whole goal area and its ground part

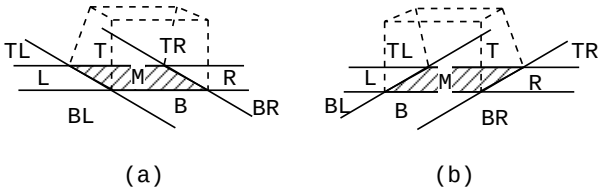


Figure 8. Relation between P and goal areas

$(Team_2, ID_2) \wedge Team_1 = Team_2 \wedge ID_1 \neq ID_2$, then we consider that event of $pass$ from P_1 to P_2 has occurred.

[Event of intercept]

If the event of $transfer$ from P_1 to P_2 occurs during I, and $P_1 = (Team_1, ID_1) \wedge P_2 = (Team_2, ID_2) \wedge Team_1 \neq Team_2$, then we consider that event of $intercept$ by P_2 against P_1 has occurred.

3.4 Relation between goal and objects

Let G be a polygon corresponding to the goal region shown in Figure 7(a), and let GG be a region of the ground part of a goal in the image projected in a 2D plane shown in Figure 7(b). We consider the relative positional relation between a player P and GG. Although GG is a parallelogram rather than a rectangle, the division of nine areas are similar to the case of a player. We represent the relation between P and GG by $r(A, I)$ where $r \in \{DC, EC, O\}, A \in S_P \cup \{M\}$ and I is an interval. Let $Tp = \{TR, T, TL\}$, $Rt = \{TR, R, BR\}$, $Lt = \{TL, T, BL\}$ and $Bt = \{BR, B, BL\}$. The fact that a player P's feet are located on GG, denoted by $in(P, GG, I)$, is described as follows.

- $O(A, I)$ and either $A = \{M\}$ or $A \cap Bt \neq \emptyset \wedge A \cap Lt \neq \emptyset$ holds, if a goal is taken from the viewpoint in the left foreground (Figure 8(a)).
- $O(A, I)$ and either $A = \{M\}$ or $A \cap Bt \neq \emptyset \wedge A \cap Rt \neq \emptyset$ holds, if a goal is taken from the viewpoint in the right foreground (Figure 8(b)).

In contrast to the players, we cannot determine the central region of the goal, since G is neither a rectangle nor a parallelogram. In this case, it is not necessary to consider the direction of the ball. If the region of the ball has a common part with G and GG during interval I, this is denoted as $in(ball, G, I)$ and $in(ball, GG, I)$, respectively.

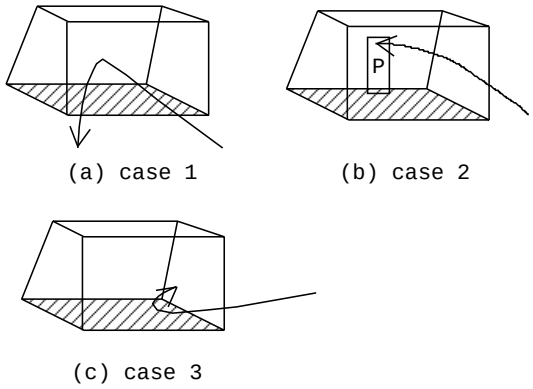


Figure 9. Shoot trajectories of success patterns

Since GG is contained in G, $in(P, GG, I) \rightarrow in(P, G, I)$ and $in(ball, GG, I) \rightarrow in(ball, G, I)$ hold.

3.5 Scored goal

Assume that a complete scene of a shot for a goal has been captured in the video without changing the scene. It means that the scene includes the beginning of an action of kicking or heading and the result of shooting, that is, a player makes contact with the ball, or the ball leaves the region G.

If a goal has been scored, then the ball must be located in G. In addition, we have to add the conditions considering the trajectory of the ball. Let I_1 be an interval in which $in(ball, G, I_1)$ holds. Let I_0 and I' be intervals such that $meets(I_0, I_1) \wedge meets(I_1, I') \wedge \neg in(ball, G, I_0) \wedge \neg in(ball, G, I')$ holds. We call the sequence $\neg in(ball, G, I_0), in(ball, G, I_1), \neg in(ball, G, I')$, *a sequence on shooting*. We have three different cases for determining the occurrence of the event of successful goal for a sequence on shooting, that are formalized as follows.

[Event of successful goal 1]

The ball passes through GG both on entering and on outgoing from G (Figure 9(a)).

(i) $starts(I_1, I_2) \wedge meets(I_2, I_3) \wedge meets(I_3, I_4)$

(ii) $in(ball, G, I_1) \wedge in(ball, GG, I_2) \wedge \neg in(ball, GG, I_3) \wedge in(ball, GG, I_4)$

[Event of successful goal 2]

A player P contacts the ball in the goal area while a ball is in G (Figure 9(b)).

(i) $finishes(I_1, I_2)$

(ii) $in(ball, G, I_1) \wedge in(P, GG, I_2) \wedge contact(P, I_2)$

[Event of successful goal 3]

A ball coming from the side of the goal hits the goal net after rebounding on GG (Figure 9(c)).

(i) $starts(I_1, I_2) \wedge meets(I_2, I_3) \wedge meets(I_3, I_4)$

(ii) $in(ball, G, I_1) \wedge \neg in(ball, GG, I_2) \wedge in(ball, GG, I_3) \wedge \neg in(ball, GG, I_4)$

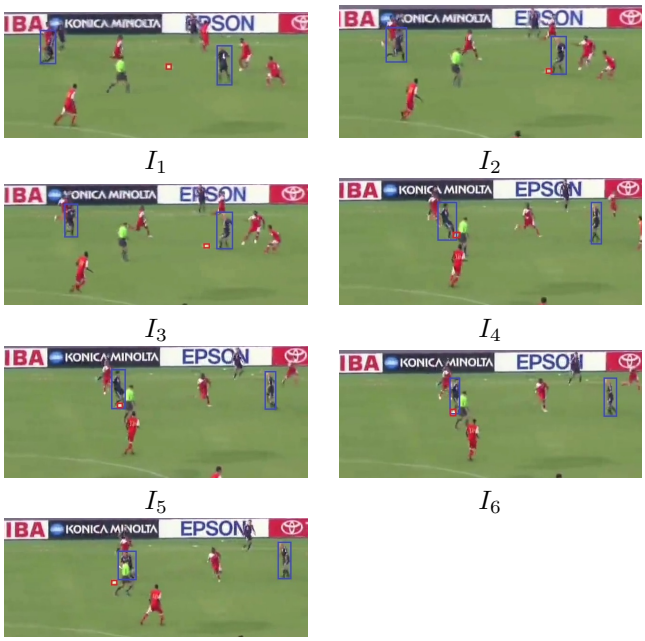


Figure 10. Time sequence of frames (passing)

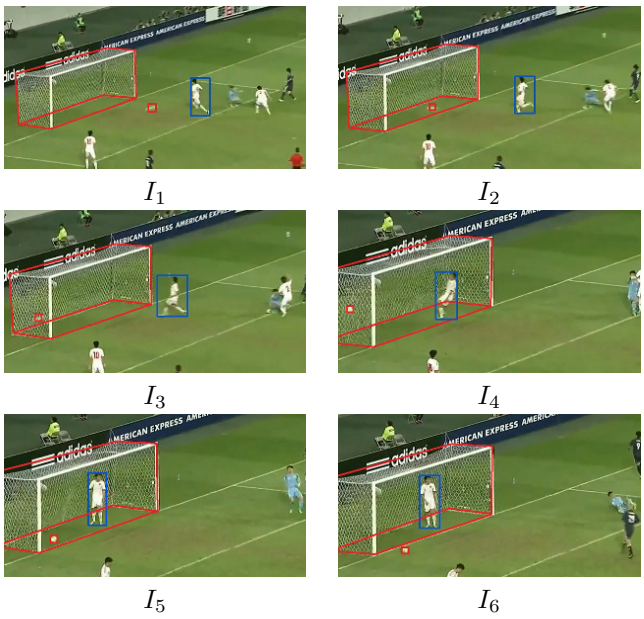


Figure 11. Time sequence of frames (shooting)

3.6 Examples

(1) pass

The following is an example of a sequence of qualitative spatio-temporal representations where $meets(I_i, I_{i+1})$ holds for each i $(1 \leq i \leq 6)$. Figure 10 shows the corresponding images. P_1 and P_2 correspond to the players enclosed by the two rectangles shown in the right and left in each frame, respectively. For example, in interval I_5, the ball is to the left of player P_1 with no intersection, and in the area occupied by P_2.

```
DC(P1,{L},I1),  DC(P2,{BR},I1)
O(P1,{L},I2), DC(P2,{BR},I2)
DC(P1,{L},I3), DC(P2,{BR},I3)
DC(P1,{L},I4), O(P2,{R},I4)
DC(P1,{L},I5), O(P2,{M},I5)
DC(P1,{L},I6), DC(P2,{B},I6)
DC(P1,{BL},I7), DC(P2,{BL},I7)
```

We determine $contact(P_1, I_2)$ from this sequence since the ball is to the left of P_1 at both I_1 and I_3, and $same(L, L)$ holds. We also determine $contact(P_2, I_4+I_5)$ since $same(BR, B)$ holds. Moreover, we find that P_1 and P_2 are on the same team, from the colors of their uniforms, for example. Therefore, we can conclude that a ball is passed from P_1 to P_2 in interval $I_2 + I_3 + I_4 + I_5$.

(2) shoot

The following illustrates another example where $meets(I_i, I_{i+1})$ holds for each i $(1 \leq i \leq 5)$. Figure 11 shows a corresponding sequence of images. Let P be a player which is enclosed by the rectangles shown in the figure. For example, in interval I_3, the ball is in the direction between L and BL of P and in the goal region.

```
DC(P,{L},I1),
DC(P,{L},I2), in(ball,GG,I2),
DC(P,{L,BL},I3), in(ball,G,I3)
DC(P,{L},I4), in(ball,G,I4),
     in(P,GG,I4)
DC(P,{BL},I5), in(ball,GG,I5),
     in(P,GG,I5)
DC(P,{BL},I6), in(P,GG,I4)
```

Let I be the interval $I_2 + I_3 + I_4 + I_5$. Then the following holds: $starts(I, I_2) \wedge meets(I_2, I_3 + I_4) \wedge meets(I_3 + I_4, I_5) \wedge in(ball, G, I) \wedge in(ball, GG, I_2) \wedge \neg in(ball, GG, I_3 + I_4) \wedge in(ball, GG, I_5)$. This satisfies the first condition of a successful shot for a goal, and we conclude that a goal has been scored in interval I.

4 Evaluation

We applied this method of event extraction to actual video data using the following process.

1. Use an image-processing tool to obtain the regions corresponding to the ball, each player, the goals, and the goal ground regions in each frame.
2. For each frame, create a qualitative representation of the target regions.
3. Take a series of continuous frames in which the same relations hold in one interval.

	correct	incorrect	undetermined
contact	93	10	17
not contact	52	5	25
total	145	15	42

Table 1. Judgment of contact (considering only positional relationship)

	correct	incorrect	undetermined
contact	106	12	2
not contact	74	8	0
total	180	20	2

Table 2. Judgment of contact (considering the holding time)

4. Extract an event from the sequence of these qualitative spatio-temporal representations.

We used ten minutes of footage from the beginning of the video, including 202 sequences on a player's contact to evaluate the judgment of event of contact. As only a few scenes in a game included the scoring of goals, we used a collection of shooting scenes in addition to the entire 90 minutes video of one game to evaluate the judgment of success of goals. We checked 29 sequences on shooting in total. Ambiguous portions were manually revised during image processing; for example, if a ball moved too rapidly to extract an object, we pointed to it manually. The frame rate was 5 frames per second. We compared the results of the proposed method to manual judgments.

Table 1 shows the results of the evaluation on the judgment of event of contact. In this table, "contact" and "not contact" denote the manual judgment, and "correct," "incorrect" and "undetermined" mean that correctly judged, incorrectly judged and undetermined by the system, respectively.

These results show that judgments were incorrect for 7.4% and could not be made for 20.8% of the sequences on a player's contact. To better judge the undetermined cases, we added another criterion. For any player P, let I be P's holding time in a sequence on P's contact. If the length of I is one frame, then P is judged as not having made contact with the ball during that sequence; if the length of I is more than two frames, then P is judged as having made contact with the ball during that sequence. The remaining cases were judged as "undetermined" in terms of whether P made contact with the ball. When we added this criterion to the judging process, the number of undetermined cases was reduced to two (Table 2). Finally, we calculated that correct judgments on whether a player made contact were provided in 89.1%.

	correct	incorrect	undetermined
success	11	0	0
fail	18	0	0
total	29	0	0

Table 3. Judgment of successful goal

The most frequent pattern, which occurred on 68 of 106 cases in which a player actually made a contact with the ball and judged correctly, is the one shown in Figure 6(a). The second most frequent pattern is the one shown in Figure 6(b), and other patterns occurred more rarely. Most of the cases in which event of contact actually occurred but was not judged as occurring involved a player kicking in the direction opposite to that of the ball's incoming direction.

In terms of shooting, all of the patterns of trajectories in the video data were judged correctly (Table 3). The most frequent successful goal pattern is depicted in Figure 9(a).

We are unable to discuss the performance of the method as image-processing techniques are out of our scope. In principle, qualitative approach should offer advantages over the usual quantitative one with respect to computation speed as the latter should handle a large amount of relatively homogeneous data.

Most studies on the extraction of events from video data have used a still camera [11] or have focused on specific events, such as a goal, free throw, or change in zoom, that can be recognized by incorporating a static object, such as a goal net, center line, and so on [12] or by including additional information such as textual data from webcasts [13]. These studies have not involved events with multiple dynamic objects, such as transferring a ball between players. Thus, the conditions considered in our study are more complex and perhaps challenging.

Our results demonstrate the effectiveness of our method; however, several problems remain.

Determinations of whether a goal has been scored involve additional factors, such as whether the offside rule was broken or a foul occurred. These should be considered, but the current method does not include them.

When a player jumps in front of the goal and blocks the shot, s/he is considered to occupy the goal ground according to the current definition. In this case, we erroneously considered a goal to have been scored.

A further problem is that broadcast video data are frequently cut when the ball reaches the goal net, making it difficult to determine whether a goal has been scored as the data representing the trajectory of the ball have been truncated. In this case, our current method cannot determine whether a shot was successful.

5 Related Works

There are several QSTR frameworks.

Hazarika et al. formalized a method for describing motion history from local surveys [14]. Weghe et al. described a trajectory-based theory to handle qualitative changes between moving objects [15]. Boxer et al. demonstrated how general physical behavior can be learned from a sequence of qualitative representations, including velocity data, using Bayesian networks [16]. However, almost all of these works handled only two-dimensional (2D) motion and the targets were not live video footage.

Santos et al. formalized data extraction from a sequence of snapshots [17, 18]. They proposed depth profile calculus (DPC) and dynamic depth profile calculus (DDPC); introduced the relation $coalescent$, which represents occlusion; and modified RCC to fit the representative image data. The main issue was to solve the problem of occlusion and how to treat a pair of objects.

Sridhar et al. described a framework for unsupervised learning of event classes from video data aimed at practical applications. In their approach, convex closures of multiple objects were extracted from video data, and relations were represented qualitatively. Learning of event classes was processed based on a probabilistic model [19]. They also proposed a more efficient method for handling noisy data [7]. However, the extracted regions were rectangular and the viewpoint of the camera was stable, which limits its application.

Recently, a new representation method called Core9 was introduced for extracting events from video data [20]. In Core9, the shape of a unit region is a rectangle, so each object is represented as a rectangle, and closure is considered for a pair of objects. The region is divided into nine cells and the relationships between the objects are represented by a matrix describing the occupation of objects for each cell. Core9 can also represent attributes including the relative size, distance, and orientation of objects, and is extended to handle leaning rectangles [21, 22]. However, because all interior angles are right angles, Core9 cannot be directly applied to event extraction of video data of football games where parallelograms or polygons are required to describe the spatial extent of objects.

6 Conclusion

We have described a qualitative representation for spatial relations of objects extracted from video data of football games, and detailed methods to determine event occurrence from the temporal sequence of these relations.

The qualitative approach differs from ordinal event extraction, which uses quantitative data. It can reduce computational complexity because it is based on symbolic computation. Moreover, it provides clear semantics for event extraction.

Our target includes scenes in which multiple objects are involved and the viewpoint of the camera can move. Moreover, we demonstrated a method for handling shapes of polygons such as a 2D projection of the goal area rather than simple rectangles.

In future work, we plan to apply this method to further data extraction and to refine the rules of event extraction. In addition, we will consider the use of information around the interval in which an event occurs. Finally, we will compare the results of learning using HMM and statistical methods, and combine this with our approach.

Acknowledgement

This work is supported by JSPS KAKENHI Grant No. 25330274.

References

[1] S. Hongeng, R. Nevatia, and F. Bremond, "Video-based event recognition: Activity representation and probabilistic recognition method," *Computer Vision and Image Understanding*, vol. 96, pp. 129–162, 2004.

[2] K. D. Forbus, "Qualitative process theory," *Artificial Intelligence*, vol. 24, no. 1-3, pp. 85–168, 1984.

[3] O. Stock, *Spatial and Temporal Reasoning*. Kluwer Academic Press, 1997.

[4] A. G. Cohn and S. M. Hazarika, "Qualitative spatial representation and reasoning: An overview.," *Fundamental Informaticae*, vol. 46, no. 1-1, pp. 1–29, 2001.

[5] G. Ligozat, *Qualitative Spatial and Temporal Reasoning*. Wiley, 2011.

[6] H. M. Dee, D. C. Hogg, and A. G. Cohn, "Scene modelling and classification using learned spatial relations," in *COSIT09*, pp. 295–311, 2009.

[7] S. Muralikshna, A. G. Cohn, and D. C. Hogg, "From video to rcc8: exploitig a distance based semantics to stabilise the interpretation of mereotopological relations," in *Proceedings of the 10th International Conference on Spatial Information Theory (COSIT11)*, Springer, 2011.

[8] http://opencv.org/

[9] D. A. Randell, Z. Cui, and A. G. Cohn, "A spatial logic based on regions and connection," in *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR92)*, pp. 165–176, 1992.

[10] J. F. Allen, "Maintaining knowledge about temporal intervals," *Communications of the ACM*, vol. 26, pp. 832–843, 1993.

[11] V. I. Morariu and L. S. Davis, "Multi-agent event recoginition in structured scenarios," in *Proceedings of Computer Vision and Pattern Recogintion (CVPR2011*, pp. 3289–3296, 2011.

[12] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Transactions on Image Processing*, vol. 12, no. 7, pp. 796–807, 2003.

[13] Y. Zhang, C. Xu, Y. Rui, J. Wang, and H. Lu, "Semantic event extratcion from basketball games using multi-modal analysis," in *IEEE International Conference on Multimedia and Expo*, pp. 2190–2193, 2007.

[14] S. M. Hazarika and A. G. Cohn, "Abducing qualitative spatio-temporal histories from partial observations," in *Proceedings of the Eighth International Conference on Principles of Knowledge Representation and Reasoning (KR02)*, pp. 14–25, 2002.

[15] N. van de Weghem, A. G. Cohn, G. de Tré, and P. de Maeyer, "A qualitative trajectory calculus as a basis for representing moving objects in geographical information systems," *Control and Cybernetics*, vol. 35, no. 1, pp. 97–119, 2006.

[16] P. A. Boxer, "Learning naive physics by visual observation: Using qualitative spatial representations and probabilistic reasoning," *International Journal of Computational Intelligence and Applications*, vol. 1, no. 3, pp. 273–285, 2001.

[17] P. E. Santos, "Reasoning about depth and motion from an observer's viewpoint," *Spatial Cognition and Computation*, vol. 7, no. 2, pp. 133–178, 2007.

[18] M. V. dos Santos, R. C. de Brito, H. H. Park, and P. Santos, "Logic-based interpretation of geometrically observable changes occurring in dynamic scenes," *Applied Intelligence*, vol. 31, pp. 161–179, 2009.

[19] S. Muralikshna, A. G. Cohn, and D. C. Hogg, "Unsupervized learning of event classes from video," in *Proceedings of the Americal National Conference on Artificial Intelligence (AAAI10)*, AAAI, 2010.

[20] A. G. Cohn, J. Renz, and M. Sridhar, "Thinking inside the box: A comprehensive spatial representation for video analysis," in *Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning (KR12)*, 2012.

[21] H. S. Sokeh, S. Gould, and J. Renz, "Efficient extraction and reresentation of spatial information from video data," in *Proceedings of the 23rd international joint conference on Artificial intelligence (IJCAI13)*, 2013.

[22] X. Ge and J. Renz, "Representation and reasoning about general solid rectangles," in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI13)*, 2013.