Reasoning about Relative Relationships in 3D Space for Objects Extracted from Dynamic Image Data

Kazuko TAKAHASHI
Kwansei Gakuin University, Japan,
email: ktaka@kwansei.ac.jp

Abstract. We describe a method of deriving relative relationships in three-dimensional space for objects extracted from video data. The method exploits information about occlusion within a framework of qualitative spatio-temporal reasoning. In particular, we are concerned with correctly deriving movements that are not entirely observable, such as an object moving through an opaque tube. Using a pair of instantaneous static images taken from two different directions, we determine the location in three-dimensional space as far as possible, then use dynamic image data from around that instant to fill in the missing pieces. Moreover, we present an envisionment that shows qualitative change of relative relationships. This method can be applied to the automatic extraction of events from video data.

Keywords: qualitative spatio-temporal reasoning, RCC, occlusion, viewpoint, envisionment

1 INTRODUCTION

Recent progress in the performance of computers provides opportunities for handling spatial data such as images or video data. Image and video data are not only frequently uploaded on the Web, but also appear as data captured by robot cameras, and require real-time analysis. We may arrange these in order, provide them with tags or keywords, or predict the behaviors of objects captured in them. In view of these possibilities, an efficient method is required for analyzing these data at an abstract level and recognizing what has happened.

This paper describes a method of deriving relative relationships in three-dimensional (3D) space between objects extracted from video data, within a Qualitative Spatio-Temporal Reasoning framework. Qualitative Spatial Reasoning (QSR) is a method that treats figures or images qualitatively, by extracting the information necessary for a user's purpose [19, 4, 12]. A system that incorporates dynamics is also called Qualitative Spatio-Temporal Reasoning (QSTR). QSTR is related to earlier research on qualitative simulation [8] aiming at the qualitative treatment of physical changes, in the sense that both handle discrete data. Cui et al. applied this idea to spatial reasoning [5]. This research was followed by a number of works on qualitative simulation [3, 10, 20]. However, most of these efforts dealt with two-dimensional (2D) objects. Some researchers have investigated the movement of solid objects that are completely filled, but few studies have focused on 3D space. When we think of objects in 3D space, visibility becomes an inevitable consideration. When an object is in the shadow of another object, it is invisible to an observer. In particular, QSTR frameworks have not discussed or formalized events in which an object passes through the inner part of another opaque object, which is a characteristic issue in 3D space. In this paper, we discuss the construction of a qualitative 3D model that includes such cases. Several researchers have used the concept of occlusion to represent the degree of visibility [16, 1, 17]. Occlusion and visibility are treated in a 2D plane [7] that is a projection of 3D space. In these works, the main objectives are axiomatization and development of a model of visibility. The authors pay little attention to mechanical reasoning algorithms, and none discuss how an occlusive relation is determined. They assume the transparency of objects; that is, they assume that a hidden part "exists but is invisible," and their reasoning is based on this assumption. However, when one object appears in an image, we do not actually know if another occluded object is also present.

In one well-known technique, a 3D model is constructed by combining multiple image data taken from different viewpoints. While numerical data such as coordinates are used in general image processing, we utilize the relationships between rectangles extracted from video data, which are closures of objects. This is a practical approach, because most image processing tools extract objects in this way. The QSTR approach is also advantageous because of its relatively small demands on memory and workspace.

In general, we cannot determine the relative relationships in 3D space between objects from one instantaneous image, because the image may include a blind spot, depending on the viewpoint and the shape of the object.

As one solution to this problem, we can predict a hidden part from continuous video data taken from a unique viewpoint [6]. For example, consider the image of 3D objects taken from a certain viewpoint shown in Figure 1. The image alone is insufficient for determining whether a part of A is hidden by B, or A is on top of B. However, if the sequence of images shown in Figure 2 is provided, we can reason that only a part of A is shown in Figure 1, and it is highly probable that A moves behind B. In contrast, if the time sequence continues to show the configuration of Figure 1, then we reason that it is highly probable that A is on top of B. However, suppose that B is a hollow tube. In this case, even if a sequence of images is given, we cannot judge whether A moves behind B, or A passes through B, if the video data is taken from a single viewpoint.

Another solution is to project 3D objects onto a 2D plane from a specific viewpoint, and then derive the positional relationship in 3D from multiple projections. Multiple image data are required to eliminate a blind spot, but it is impossible to choose such a viewpoint that completely eliminates the blind spot for some types of objects. For example, assume that a tiny ball is moving around a big ball. In this case, it is impossible to find a stable viewpoint from which the

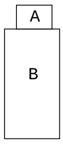


Figure 1. Objects A and B at an instant

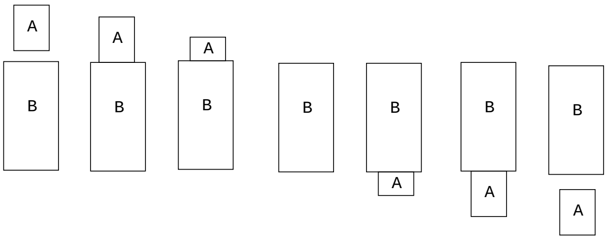


Figure 2. Objects A and B in a time sequence

tiny ball can be observed at any instant. Therefore, to create a 3D model from 2D data, we need a sequence of image data taken from multiple viewpoints. The greater the number of viewpoints is, the more reliable is the reasoning.

In this paper, we describe the reasoning about relative relationships of objects based on data about rectangles extracted from videos taken from two different viewpoints. In particular, for a hollow tubular object, we show a method for identifying an event in which another object passes through it. We also determine whether the tube has a cap from its relative relation with another moving object. Moreover, we demonstrate that we can reason the direction of movement from a sequence of image data taken from a unique viewpoint, and present an envisionment.

This paper is organized as follows. In Section 2, we describe RCC, which is the QSR framework on which our work is based. In Section 3, we discuss our qualitative model. In Section 4, we explain the technique for constructing a 3D model from 2D data. In Section 5, we demonstrate the reasoning of relations in 3D space. In Section 6, we compare our work with related research. Finally, in Section 7, we present our conclusions.

2 Region Connection Calculus

Region Connection Calculus (RCC) is one of the representatives of QSR frameworks [15], lots of QSR systems based on. We also construct a model based on RCC. In RCC, spatial data are represented as relative positional relationships of regions. In RCC8, which is the most popular among several RCC systems, only the connections of regions are considered, and other information is ignored.

Figure 3 shows the eight primitives in RCC8: DC (disconnected), EC (external connection), PO (partial overlap), EQ (equal), TPP (tangential proper part), NTPP (non-tangential proper part); TPPi and NTPPi are inverse relations of TPP and NTPP, respectively. These primitives are jointly exhaustive and pairwise disjoint (JEPD). In this figure, only states connected by an edge can transit directly from one another. For example, if a pair of regions is in the PO relation, then it does not change to the state DC without passing through the state EC.

RCC itself is free from the concept of visibility.

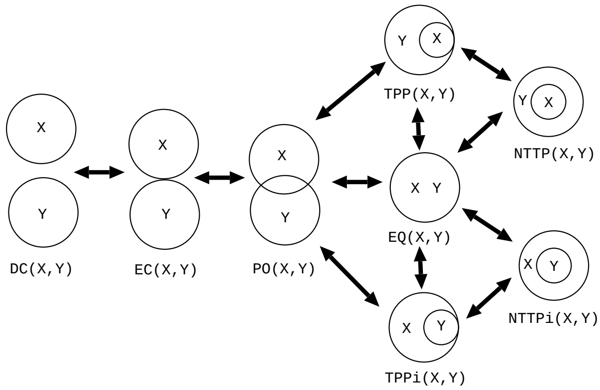


Figure 3. Fundamental relationships of RCC8

3 DESCRIPTION OF THE QUALITATIVE MODEL

3.1 Target objects

Our 3D target objects are classified into two types: *solid objects* and *tube objects*. A solid object is one that is completely filled. A tube object is one that is hollow, allowing another object to pass through or be wholly or a partly contained within it. We assume that an object has no holes and no dents, and that the boundary lines of objects are straight. An object can move, but its inherent shape or size is stable. We also assume that no object splits, is united with another object, is newly created, or becomes extinct. An object that does not move is called *a static object*, and an object that moves is called *a dynamic object*.

The minimum convex polyhedron containing an object is called *the closure* of the object. The closure contains the object itself as its boundary. If an object is twisted or is a tube, then the object and its closure do not coincide. For such an object, we regard the closure as a region occupied by the object. Therefore, when an object X is entirely inside a tube object Y, their RCC8 relationship is NTPP(X,Y) or TPP(X,Y) (Figure 4).

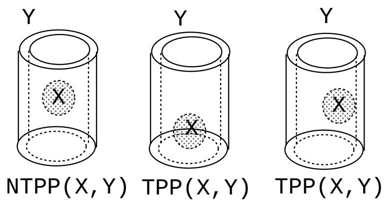


Figure 4. TPP and NTPP relations w.r.t. a tube object

We introduce the concepts of body and image for an object. Given an object x, $bd(x,t)$ denotes its body at an instant t, and refers to the region x actually occupies in 3D space at that instant, while $im(x, v, t)$ denotes its image from a viewpoint v at an instant t, and refers to its projection in the designated direction.

We assume that each $im(x, v, t)$ is a rectangle, because the object data extracted from the image data is rectangular in shape. When the images of two objects are externally connected by a point or a line, their relation is EC in both cases, and when one is a tangential proper part via a point or a line, their relation is TPP in both cases. Moreover, when an object x is partially hidden by another object y, and its visible part in the image is concave, the relation of $im(x, v, t)$ and $im(y, v, t)$ is not EC, but PO. This fact results in the indeterminacy of the question of which object is in the foreground with regard to a viewpoint if their relation is PO (Figure 5).

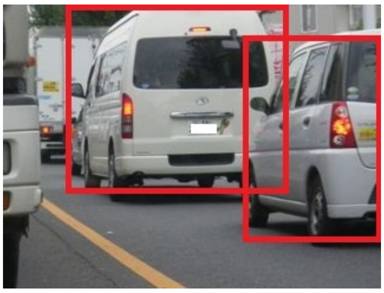


Figure 5. Indeterminacy of the layering of objects in a PO relation

Because we do not assume the transparency of objects, and we have no information about the hidden part, RCC is not suitable for representing invisibility. Therefore, we introduce new predicates to represent a situation in which only one object is observed. $Z(x, y)$ indicates that only y is observed, and $Zi(x, y)$ indicates that only x is observed.

To simplify the problem, as a first step, we discuss the relation between two objects x and y that satisfy the following conditions.

- x is a dynamic solid object that is smaller than y.
- y is a static object whose shape is either an n-prism, a pole, or an L-shaped column (Figure 6).
- The relative size of the objects in each image from any viewpoint is stable at every instant.

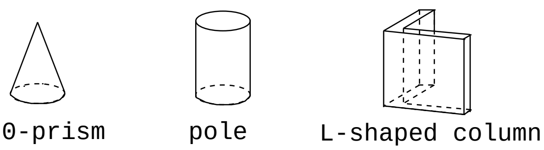


Figure 6. Examples of objects admitted as y

The purpose of the second condition is to reduce the number of blind spots. Note that a ball, for example, cannot be allowed. This constraint ensures that when an object is not observed in an image taken from some viewpoint, this does not mean that the object is located in a blind spot, but that it is located in back of another object.

We denote by $\mathbf{R_2}$ and $\mathbf{R_3}$ the sets of relationships of images in the 2D plane and bodies in 3D space, respectively. Let $\mathcal{R}_2$ be $\{DC, EC, PO, TPP, NTPP, Z\}$. These relations are jointly exhaustive and pairwise disjoint. For $\mathbf{R_2} \in \mathcal{R}_2$, we write $\mathbf{R_2}(x, y, v, t)$ instead of $\mathbf{R_2}(im(x, v, t), im(y, v, t))$. Let $\mathcal{R}_3$ be $\{\mathbf{DC}, \mathbf{EC}, \mathbf{PO}, \mathbf{TPP}, \mathbf{NTPP}\}$. These relations are jointly exhaustive and pairwise disjoint. For $\mathbf{R_3} \in \mathcal{R}_3$, we write $\mathbf{R_3}(x, y, t)$ instead of $\mathbf{R_3}(bd(x, t), bd(y, t))$. Moreover, we can omit t when the meaning is clear.

Due to the above constraints, $EQ, TPPi, NTPPi$, and Zi never appear. The following axiom reflects this property.

[Axiom1] $\forall v(\exists t(TPP(x, y, v, t) \vee NTPP(x, y, v, t)) \longrightarrow \forall t'(\neg TPPi(x, y, v, t') \wedge \neg NTPTi(x, y, v, t') \wedge \neg EQ(x, y, v, t') \wedge \neg Zi(x, y, v, t'))$

3.2 Viewpoint

For a space constituted by the above x and y, two viewpoints v_u and v_s are specified as follows: v_u points in the direction of y's base. v_s points in the direction of y's side. v_u and v_s are called the *upper viewpoint* and *side viewpoint*, respectively.

We introduce the predicates $fore$ and $back$ indicating which object is nearer from a specified viewpoint.

$fore(x, y, v, t)$: x is in the foreground of y, namely, x is nearer to v than y at the instant t.

$back(x, y, v, t)$: x is in the background of y, namely, x is farther from v than y at the instant t.

Formally, they are defined as follows. Let $dist(p, q)$ denote the distance between the points p and q in 3D space, and let p_x and p_y denote points in $bd(x, t)$ and $bd(y, t)$, respectively. Then

$fore(x, y, v, t) =_{def} \forall p_x \forall p_y.(dist(p_x, v) \leq dist(p_y, v))$

$back(x, y, v, t) =_{def} \forall p_x \forall p_y.(dist(p_x, v) \geq dist(p_y, v))$

The following axiom indicates the continuity of a transition of relative positional relations. It specifies that if the foreground/background relation of two objects is changed, then there exists an instant in which a change of foreground and background occurs.

[Axiom2] $\forall v((back(x, y, v, t_1) \wedge fore(x, y, v, t_2)) \longrightarrow \exists t((t_1 \leq t \leq t_2) \wedge PO(x, y, v, t)))$

4 MODELING AND REASONING

4.1 Construction of a qualitative 3D model

We represent the relations of rectangles extracted from an image as $\mathbf{R_2}$ relations. Then we construct a qualitative 3D model from this set of relations via the following process.

First, we derive the $\mathbf{R_3}$ relation of the bodies of two objects from a single image based on a single viewpoint. The relation is uniquely determined in some cases. If it is not determined, then we derive it from a pair of images taken at the same instant. If we still cannot determine the relation, we check the dynamic change from a specific viewpoint.

(1) Derivation from a single image based on a single viewpoint

If two objects are observed to be disconnected from a certain viewpoint, then they are disconnected in 3D space.

[Rule 1] $DC(x, y, v_s) \vee DC(x, y, v_u) \longrightarrow \mathbf{DC}(x, y)$

(2) Derivation from a pair of images taken at the same instant

If two objects are observed to be externally connected from a side viewpoint, and not disconnected from an upper viewpoint, then they are externally connected in 3D space.

[Rule 2] $EC(x, y, v_s) \wedge \neg DC(x, y, v_u) \longrightarrow \mathbf{EC}(x, y)$

If two objects are observed to be externally connected from a side viewpoint, and partially overlapped, or only one object is observed from an upper viewpoint, then they are externally connected in 3D space.
[Rule 3] $EC(x, y, v_s) \wedge (PO(x, y, v_u) \vee Z(x, y, v_u)) \longrightarrow \mathbf{EC}(x, y)$

If only one object is observed from a side viewpoint, and the other is observed to be a tangential proper part from an upper viewpoint, then it is a tangential proper part of the other object in 3D space.
[Rule 4] $(Z(x, y, v_s) \wedge TPP(x, y, v_u)) \longrightarrow \mathbf{TPP}(x, y)$

(3) Focusing relations in 3D space

In the following cases, we cannot determine a unique relation in 3D space, but can narrow the result to two possible relations.

If two objects are observed to be externally connected from an upper viewpoint, and one is a (non-)tangentially proper part from a side viewpoint, then they are externally connected or partially overlapped in 3D space.
[Rule 5] $EC(x, y, v_s) \wedge (TPP(x, y, v_u) \vee NTPP(x, y, v_u) \vee Z(x, y, v_u)) \longrightarrow \mathbf{EC}(x, y) \vee \mathbf{PO}(x, y)$

If only one object is observed from a side viewpoint, and only one object is observed, or the other is a non-tangentially proper part from an upper viewpoint, then it is a (non-)tangentially proper part in 3D space. In this case, we can refine our judgment no further without assuming the transparency of y.
[Rule 6] $(Z(x, y, v_s) \wedge (NTPP(x, y, v_u) \vee Z(x, y, v_u)) \longrightarrow \mathbf{TPP}(x, y) \vee \mathbf{NTPP}(x, y)$

Table 1 lists the rules used to derive a relation in 3D space from 2D data. In this table, "-" means "impossible".

Table 1. Deriving relations in 3D space from 2D data

$v_s \backslash v_u$	DC	EC	PO	TPP	NTPP	Z
DC	DC	DC	DC	DC	DC	DC
EC	DC	EC	EC	EC	EC	EC/PO
PO	DC	EC	-	-	-	-
TPP	DC	EC	-	-	-	-
NTPP	DC	EC	-	-	-	-
Z	DC	EC	-	TPP	TPP/NTPP	TPP/NTPP

If y is known to be a solid object, either $\mathbf{DC}$ or $\mathbf{EC}$ holds in 3D space. Therefore, Rule 5 is replaced by the following Rule 5', where $solid(y)$ indicates that y is a solid object.
[Rule 5'] $solid(y) \wedge EC(x, y, v_s) \longrightarrow \mathbf{EC}(x, y)$

If y is not known to be a solid object, we can narrow the result by using data from a time t' which is different from t. This is the situation in which x moves between t and t' and part of x is included in y at time t (Figure 7).

(4) Derivation from a dynamic change of a specific viewpoint
[Rule 7] $EC(x, y, v_s, t) \wedge (TPP(x, y, v_u, t) \vee NTPP(x, y, v_u, t) \vee Z(x, y, v_u, t)) \wedge \exists t'(\ (t' \neq t) \wedge TPP(im(x, v_s, t), im(x, v_s, t'))\ ) \longrightarrow \mathbf{PO}(x, y, t)$

(5) More intelligent derivation

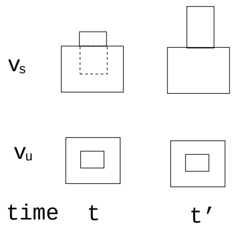


Figure 7. A case in which it is determined that PO holds at an instant t

The example shown in Figure 8(a)(b) is a more complex case.

In the image taken at time t, the three objects x_1, x_2 and y are observed, and have the following relationships:

$EC(x_1, y, v_s, t) \wedge EC(x_2, y, v_s, t) \wedge DC(x_1, x_2, v_s, t) \wedge TPP(x_1, y, v_u, t) \wedge Z(x_2, y, v_u, t) \wedge Z(x_2, x_1, v_u, t)$

Note that x_2 is not hidden by y in the image taken from the upper viewpoint, but the relation is represented as $Z(x_2, y, v_u, t)$.

In the image taken at time t', the following relationship holds:

$EC(x, y, v_s, t') \wedge TPP(x, y, v_u, t')$

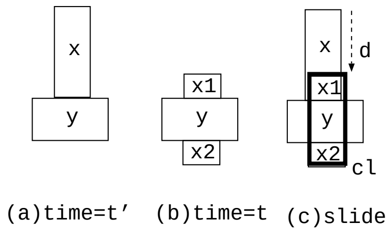


Figure 8. Objects separately observed from v_s

When such 2D data are given, we cannot determine whether both x_1 and x_2 are parts of the same object x, or are totally different objects. According to Rule 7, $\mathbf{PO}(x, y, t)$ holds, and x_1 is considered to be a part of x. However, we are uncertain about x_2; this is due to the fact that although the number of objects is stable, one of them may be invisible (Figure 9). Therefore, the case shown in Figure 9 is also possible.

We define a new function $slide$ to solve this problem.

$slide(im(x, v, t), d)$ is the function that returns the data obtained by transferring the image of x from v at the instant t according to the specified vector d (Figure 8(c)). Let cl be the closure of $im(x_1, v_s, t)$ and $im(x_2, v_s, t)$, which is shown in the boldface frame of Figure 8(c). If $slide(im(x, v_s, t), d)$ coincides with cl, then x_2 is considered to be a part of x.

Therefore, we have the following rule:

[Rule 8] $EC(x_1, y, v_s, t) \wedge EC(x_2, y, v_s, t) \wedge \exists t'(\ (t' \neq t) \wedge DC(x, y, v, t') \wedge TPP(im(x_1, v_s, t), im(x, v_s, t')) \wedge \exists d\ EQ(slide(im(x, v_s, t'), d), cl)\ ) \longrightarrow \mathbf{PO}(x, y, t)$

If this condition does not hold, then x_1 and x_2 are different objects.

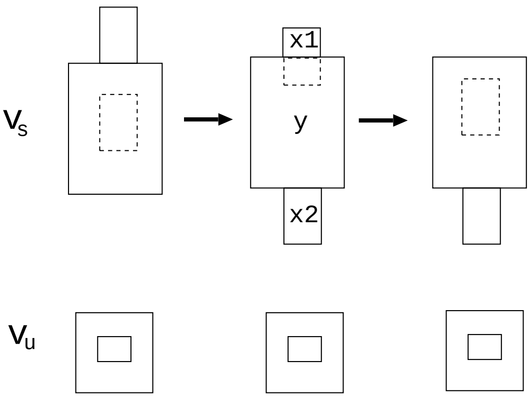
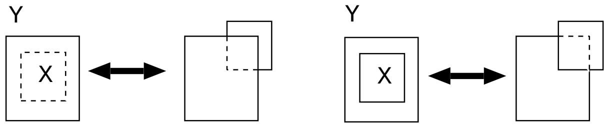


Figure 9. A case of different objects



(a) back(x,y,v,t) (b) fore(x,y,v,t)

Figure 10. Determining the foreground/background

4.2 Determining a foreground/background relationship

In addition to RCC relations in 3D space, we can determine which object is located in the foreground from a specific viewpoint.

Given a tube object y, $has\_cap(y, v)$ indicates that it has a cap in the v direction, and $has\_bot(y, v)$ indicates that it has a cap in the inverse direction to v.

We specify the rules for deriving the foreground and background relation of two objects with regard to a given viewpoint.

If only y is observed from the side viewpoint, then x is in the background of y.

[Rule 9] $Z(x, y, v_s) \longrightarrow back(x, y, v_s, t)$

If x is a (non-)tangentially proper part, then x is in the foreground of y.

[Rule 10] $(TPP(x, y, v_s) \vee NTPP(x, y, v_s)) \longrightarrow fore(x, y, v_s, t)$

If only y is observed from the upper viewpoint, then x is in the background of y.

[Rule 11] $Z(x, y, v_u) \longrightarrow back(x, y, v_u, t)$

If x is a (non-)tangentially proper part and y has a cap, then x is in the foreground of y; otherwise, it is not determined.

[Rule 12] $(TPP(x, y, v_u) \vee NTPP(x, y, v_u)) \wedge has\_cap(y, v_u) \longrightarrow fore(x, y, v_u, t)$

When $PO(x, y, v, t)$ holds, we cannot determine the foreground and background relation of two objects at t from the given data. In this case, we can derive the relation by using data from a subsequent or previous instant (Figure 10).

[Rule 13] $\exists t'(\ (t' \neq t) \wedge Z(x, y, v, t') \wedge PO(x, y, v, t)\ ) \longrightarrow back(x, y, v, t') \wedge back(x, y, v, t)$

[Rule14] $\exists t'(\ (t' \neq t)) \wedge TPP(x, y, v, t') \wedge PO(x, y, v, t)\ ) \longrightarrow fore(x, y, v, t') \wedge fore(x, y, v, t)$

4.3 Determining the direction of motion

So far, we have assumed that relative size of images of objects is invariant from both viewpoints at any time. If an object moves toward a viewpoint or away from it, this assumption does not hold. In this subsection, we allow the relative size of objects to vary, and discuss the derivation of the direction of motion from the transition of $\mathbf{R_2}$. By comparing the sizes at instants t and t', the directions of motion from t to t' are classified into three types: $sameDistMove$, $goFar$, and $comeNear$.

If x moves while maintaining the same distance from a viewpoint v, then the size of $im(x, v, t)$ is invariant. This type of movement is referred to as $sameDistMove$.

[Rule 15] $\exists d \exists t'(\ (t \neq t') \wedge EQ(slide(im(x, v, t'), d), im(x, v, t))\ ) \longrightarrow sameDistMove(x, t, t')$

If x moves toward a viewpoint or away from it, then the size of $im(x, v, t)$ varies. This type of movement is either $goFar$ (departing from v) or $comeNear$ (approaching v), depending on the situation.

[Rule 16] $\exists t'(\ (t < t') \wedge (TPP(im(x, v, t), im(x, v, t')) \vee NTPP(im(x, v, t), im(x, v, t')))\ ) \longrightarrow goFar(x, t, t')$

[Rule 17] $\exists t'(\ (t > t') \wedge (TPP(im(x, v, t), im(x, v, t')) \vee NTPP(im(x, v, t), im(x, v, t')))\ ) \longrightarrow comeNear(x, t, t')$

5 REASONING FOR A 3D MODEL

We describe the reasoning for inferring relations in 3D space.

5.1 Reasoning for a tube object

We determine whether a tube object has a cap or a bottom from the transition of relations in 3D space.

[Rule 18] $\exists t'(\ (t > t') \wedge \mathbf{EC}(x, y, t) \wedge \mathbf{PO}(x, y, t')\ ) \longrightarrow \neg has\_cap(y, v_u)$

[Rule 19] $\exists t'(\ (t < t') \wedge \mathbf{EC}(x, y, t) \wedge \mathbf{PO}(x, y, t')\ ) \longrightarrow \neg has\_bot(y, v_u)$

Moreover, if only one object is observable from the upper viewpoint, then we can infer that y has either a cap or a bottom.

[Rule 20] $Z(x, y, v_u) \longrightarrow has\_cap(y, v_u) \vee has\_bot(y, v_u)$

5.2 Event retrieval

The occurrence of an event can be retrieved from a sequence of 3D relations $\mathbf{R_3}(x, y, t_0), \ldots, \mathbf{R_3}(x, y, t_n)$ $(n \geq 1)$, where t_{i+1} is the next time instant after t_i, denoted by $next(t_i)$, for each i $(0 \leq i < n - 1)$.

We present the definitions of several events.

[Def 1] x enters y.
$enter(x, y) =_{def} \mathbf{EC}(x, y, t) \wedge \mathbf{PO}(x, y, next(t)) \wedge \mathbf{TPP}(x, y, next(next(t)))$

[Def 2] x exits y.
$exit(x, y) =_{def} \mathbf{TPP}(x, y, t) \wedge \mathbf{PO}(x, y, next(t)) \wedge \mathbf{EC}(x, y, next(next(t)\ )$

[Def3] x passes through y if x enters y and successively exits y.
$pathThrough(x, y) =_{def} \mathbf{EC}(x, y, t) \wedge \mathbf{PO}(x, y, next(t)) \wedge$

$\mathbf{TPP}(x, y, next(next(t))) \wedge \mathbf{PO}(x, y, next(next(next(t)))) \wedge \mathbf{EC}(x, y, next(next(next(next(t)))))$

[Def4] x fits inside y if the positional relationship is unchanged after entering.
$fitIn(x, y) =_{def} \mathbf{EC}(x, y, t) \wedge \mathbf{PO}(x, y, next(t)) \wedge \mathbf{TPP}(x, y, next(next(t))) \wedge \mathbf{PO}(x, y, next(next(next(t)))) \wedge \mathbf{PO}(x, y, next(next(next(next(t)))))$

5.3 Envisionment

The transition of $\mathbf{R_3}$ relations in 3D space follows Figure 3. However, the transition of $\mathbf{R_2}$ differs from the usual transitions, because we do not assume the transparency of objects.

Figure 11 and Figure 12 show the transitions of $\mathbf{R_2}$ from the viewpoint v, assuming that the relative size of objects from v is unvarying and varying, respectively.

Figure 11 shows the case in which $im(x, v, t)$ is always smaller than $im(y, v, t)$.

In Figure 12, transition is possible if x moves while maintaining the same distance from the viewpoint v, or changes its distance from v. For example, transition from PO to TPP is possible if the motion of x is $sameDistMove$, or x moves away from v and approaches y.

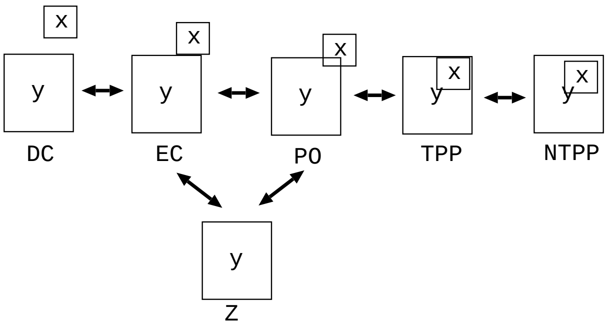


Figure 11. Envisionment: the relative size of the objects is unvarying

Assume that the size of a moving object x is bigger than the entrance of a static tube object y. In this case, x never enters y. Therefore either $\mathbf{DC}$ or $\mathbf{EC}$ holds in 3D space. As for the 2D plane, TPPi/NTPPi can hold instead of TPP/NTPP. Therefore, the transition graph for this case is obtained by replacing these parts.

6 RELATED WORK

The reasoning used to derive the relationships of moving solid objects has been well studied, both qualitatively and quantitatively. In these works, the foreground and background relations between objects are determined mostly from the complement of the continuous data. However, these works do not consider tube type objects or objects with containers. Galton formalized various types of objects, including containers, and discussed their properties [9]. However, he did not discuss the use of practical video data. In contrast, our objective is the automatic extraction of data and formalization of the method.

A number of works have focused on qualitative simulation. Bennett et al. explored the expressive power of region-based geometry

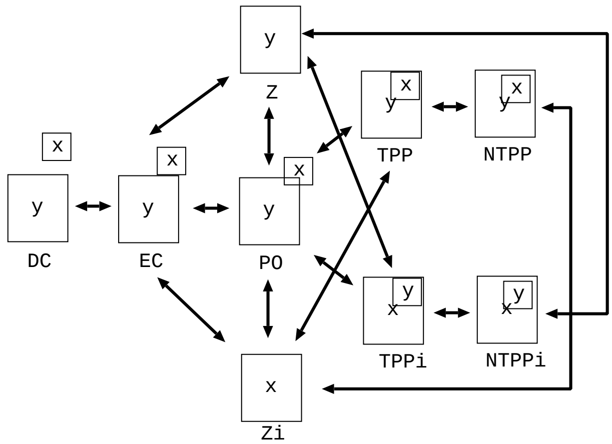


Figure 12. Envisionment: the relative size of the objects is varying

[2]. They formalized various types of movements in a qualitative manner. Hazarika et al. formalized the abduction of a motion history from local surveys [10]. Weghe et al. presented a trajectory-based theory to handle qualitative changes between moving objects [20]. Boxer et al. demonstrated how general physical behaviors can be learned from a sequence of qualitative representations with direction and velocity, using Bayesian networks [3]. Almost all of these works involve only 2D motions, and issues that arise when formalizing 3D motions, such as the occlusion problem or the tube-passage problem are not discussed.

Randell et al. presented an interesting work on occlusion [16]. They proposed ROC20, a refined RCC8 system that describes relative relationships of objects, including tube type objects, from a single viewpoint. They discussed the relationships between changes of viewpoint and changes of relationship. However, their work is based on the idea that the reasoner knows the locations of the target objects a priori, and investigates the relationships in the 2D plane for the corresponding 3D data. Their objective is to formally represent a situation using QSR. Unlike their procedure, we begin with rectangles extracted from the video data, from which we cannot know the state of the background of an object, with the objective of determining 3D relative positional relations between objects.

Santos et al. formalized abduction from a sequence of snapshots [18, 6]. They proposed Depth Profile Calculus (DPC) and Dynamic Depth Profile Calculus (DDPC). They introduced the relation $coalescent$, which represents occlusion, and modified RCC to fit the representative image data. They discussed the predication of invisible parts. Their basic idea is similar to ours, but their approach is different. They use an image obtained from a single viewpoint, and do not consider the tube-passage problem. They represent image data using the three elements of distance, size, and depth, and retrieve events in 3D space from temporal sequences of these data. In contrast, we use image data obtained from two different viewpoints in a single instant, as well as video data from around that instant, and derive RCC relations in 3D space from these. Moreover, we handle a tube object by using image data from a pair of viewpoints.

Fogliaroni et al. investigated the relationship between viewpoints and blind spots, and demonstrated the reasoning in a QSR framework. They applied their technique to localization and navigation [7].

Sridhar et al. presented a framework for unsupervised learning of event classes from video data, aimed at practical application. In their approach, convex closures of multiple objects are extracted from video data, and their relations are represented qualitatively. The learning of event classes is processed based on a probabilistic model [13]. They also proposed a more efficient method for handling noisy data [14]. They regard video data as a projection of 3D objects onto a 2D plane. In contrast, we treat 3D objects as 3D entities, rather than as projections.

In the research areas of the visual language or image processing, researchers have proposed methods that facilitate RCC over 3D [11, 1, 17]. These are based on projection onto the xyz-axes, and also assume sufficient information about the location in 3D space. However the formalization in these works is not sufficient, and they do not refer to the envisionment.

7 CONCLUSION

We have described a method of deriving relative relationships in three dimensional space for objects extracted from video data in a QSTR framework. We use image data obtained from two different viewpoints at a single instant, as well as video data from around that instant.

The proposed method offers the following advantages.

1. The relative relationships of objects in 3D space can be derived without assuming the transparency of objects.
2. Tube type objects and events related to such objects can be handled.
3. The method can be applied to the automatic extraction of events from video data.

As this is a first step toward automated qualitative recognition of relations between objects in 3D space from dynamic image data, we imposed several restrictions on the objects to be handled. In the future, we will generalize this method, investigate the properties more deeply, and perform experiments using actual video data.

ACKNOWLEDGEMENTS

This research is supported by KAKENHI222500141.

REFERENCES

[1] J. Albath, J. L. Leopold, C. L. Sabharwal, and K. Perry, 'Efficient reasoning with rcc-3d', in *Proceedings of Knowledge Science, Engineering and Management 2010 (KSEM2010)*, number 6291 in LNAI, pp. 470–481. Springer, (2010).

[2] B. Bennett, A. G. Cohn, P. Torrini, and S. M. Hazarika, 'Describing rigid body motions in a qualitative theory of spatial regions', in *Proceedings of AAAI-2000*, pp. 503–509, (2000).

[3] P. A. Boxer, 'Learning naive physics by visual observation: Using qualitative spatial representations and probabilistic reasoning', *International Journal of Computational Intelligence and Applications*, **1**(3), 273–285, (2001).

[4] A. G. Cohn and S. M. Hazarika, 'Qualitative spatial representation and reasoning: An overview.', *Fundamental Informaticae*, **46**(1-1), 1–29, (2001).

[5] Z. Cui, A. G. Cohn, and D. A. Randell, 'Qualitative simulation based on a logical formalism of space and time', in *Proceedings of AAAI-92*, pp. 679–684, (1992).

[6] 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*, **31**, 161–179, (2009).

[7] P. Fogliaroni, J. Wallgrün, E. Clementini, F. Tarquini, and D. Wolter, 'A qualitative approach to localization and navigation based on visibility information', in *COSIT09*, pp. 312–329, (2009).

[8] K. D. Forbus, 'Qualitative process theory', *Artificial Intelligence*, **24**(1-3), 85–168, (1984).

[9] A. Galton, 'The formalities of affordance', in *Proceedings of ECAI2010 Workshop on Spatio-Temporal Dynamics*, pp. 1–6, (2010).

[10] 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).

[11] C. Li, J. Chao, C. Yin, and L. Ma, 'Qualitative spatial representation and reasoning in 3d space', in *Proceedings of 2009 International Conferences on Intelligent Computation Technology and Automation(ICICTA)*, pp. 653–657. IEEE, (2009).

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

[13] 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).

[14] 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).

[15] D. Randell, Z. Cui, and A. Cohn., 'A spatial logic based on regions and connection', in *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning (KR92)*, pp. 165–176. Morgan Kaufmann Publishers, (1992).

[16] D. Randell, M. Wikowski, and M. Shanahan, 'From images to bodies: Modelling and exploiting spatial occlusion and motion parallex', in *Proceedings of the 17th international joint conference on Artificial intelligence (IJCAI01)*, pp. 57–63, (2001).

[17] C. Sabharwal, J. L. Leopold, and N. Eloe, 'A more expressive 3d region connection calculus', in *Proc. of the 2011 International Workshop on Visual Languages and Computing*, pp. 307–311, (2011).

[18] P. E. Santos, 'Reasoning about depth and motion from an observer's viewpoint', *Spatial Cognition and Computation*, **7**(2), 133–178, (2007).

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

[20] 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*, **35**(1), 97–119, (2006).