Qualitative Reconstruction and Update of an Object Constellation

H. Joe Steinhauer[1]

Abstract. We provide a technique for describing, reconstructing and updating an object constellation of moving objects. The relations between the constituent objects, in particular axis-parallel and diagonal relations, are verbally expressed using the double cross method for qualitatively characterizing relations between pairs of objects. The same underlying representation is used to reconstruct the constellation from the given description.

1 Motivation

Sections 1, 2, and 3 of this paper state the general motivation for the work from an applicational point of view. The reader who wishes to focus on the technical contents is invited to proceed directly to section 4. The first part of the research presented here has previously been described in [16]. However the understanding of the fundamental ideas presented there are necessary to understand the processes of describing and reconstructing a constellation of objects with orientations, and to understand the updating of a scene of moving objects.

Imagine a disaster area, e.g. a rapidly spreading forest fire, a flood or an area after an earthquake. Rescue experts are busy with organizing the rescue teams and equipment. To do their job as well as possible it would be advantageous if they would know what the disaster area looks like. Where are people trapped? Where is the flood rising and where sinking? In which direction does the fire spread?

In contaminated areas where it is extremely difficult for humans to check out the disaster terrain autonomous agents like all-terrain vehicles and helicopters (e.g. the WITAS helicopter [2], [10], [11]) can take over the job of exploring the area and distributing survival kits to trapped people.

Normally human rescue experts are only experienced in their specific rescue field. It would be time consuming and a source of error if the experts needed to learn how to interpret all the data that the autonomous systems collect. It would also be of disadvantage if additional special trained people were needed to translate the data collected from the autonomous systems into natural language understandable by the rescue experts.

The scenarios above indicate that the communication between the human experts and the autonomous system should take place directly in natural language. The autonomous system itself will interpret the collected data and translate its findings into human natural language. The system also needs to understand natural language input about what to do next and to answer further questions whether some specific information is also implicit in the collected data but has not been uttered yet.

[1] Department of Computer Science, Linköping University, Sweden, email: joest@ida.liu.se

In this paper we describe a process for an autonomous system, called the observer, e.g. a helicopter, that is flying over an area with moving objects and reports what is going on in this area in natural language. That means that it first describes the scene, its participants and their orientations and then updates the objects' relative positions whenever such an update is needed.

We describe further the process of another autonomous system, called the listener, that is able to reconstruct the scene from the earlier produced natural language description of the scene.

2 Qualitative aspects of natural language communication

Even though natural language is ambiguous it is a very good communication tool. Humans are highly skilled in describing scenes and situations in natural language. As we do not have any other possibility to pass on our sensor data we need to use language as long as we do not choose to draw a picture. But drawing a picture takes time, mistakes are difficult to correct afterwards and additional tools like paper and pen are needed. To transmit a picture over a distance more resources are needed than to transmit natural language which can be done by phone, walkie-talkie, sending in light Morse, or even by shouting.

When we describe a scene we provide a description of an already interpreted image. Before we utter a description we recognize objects (vehicles, people, plants), classify them (cars, trucks, pedestrians, trees), maybe group them (a convoy of trucks, a queue of cars, a crowd of people, a forrest) and we very often abstract from details (how many people are in the crowd? How many trucks form the convoy? Are all trucks of the same size? And how many trees are in the forest?).

We do not explain exactly where the objects are in world coordinates and we rarely attach a coordinate system to a room in which the party took place that we want to describe to our friend. We instead use the objects' relative positions to each other to describe what was where. We say that the pen lies on the desk, or that my car is parked to the left of yours.

Furthermore we are not so good at giving exact quantitative information of measurable things either. We normally prefer to use qualitative information and say that some object is *bigger* than the other or *further* away.

In this paper we argue that using qualitative relational information for describing an object constellation in natural language and updating this constellation whenever at least one of the qualitative relations between the objects changes is sufficient to communicate what the scene looks like and how it develops. No further quantitative information is needed for the purpose of submitting a description that

would be similar to a human observer's natural language description.

3 "Mental images"

When two computers exchange information and as long as speed problems are not of concern they can use a data link and could pass on all the original data, so that the other computer gets exactly the same information as the first one.

Normally though a computer program is expected to interpret the collected data before it passes it on to another application. An expert system that might be used to interpret the scene that the autonomous helicopter observes would hardly work on the pixel information the helicopter's camera collects but would instead expect a vision system to recognize the objects and to pass on the semantic information of the scene.

Especially if one of the communication parties is a human being it is not at all practical to just pass on the collected data. A human listener would have great difficulties to finally extract the implicit information from the input and a human observer barely has the possibility to provide an uninterpreted description in this way. The input data that reaches his retina and are processed in his brain are not accessible for him. The result of this processes is an interpreted picture in his mind which is a mental image of what he saw. The best he can do is to provide a description of this mental image.

The term *mental images* is frequently used in the literature of natural language systems, dialog systems, and vision systems. Ideally a vision system should describe the scene it sees in such a way that the human listener's mental image of the scene becomes identical with what it would be if the listener would watch the scene himself [1], [6], [12].

To do so the observer system first uses image processing and object recognition mechanisms to recognize objects and to classify them. How this is done is not within the scope of our work.

After all objects are recognized and classified the representation of this could be seen as a "mental image" of the scene, in the sense that this data collection contains an interpretation of the image. The alternative of this interpretation would be an uninterpreted description that for example just contains the color and hue value of each pixel in the scene.

4 The communication model

The motivation of our work is the alleviation of human-machine communication. Imagine the autonomous helicopter, mentioned before, that is flying over a disaster area and reports what is going on to a human listener. In this case it would be of advantage for the listener if the helicopter would use natural language.

On the other hand you might think of an expert system that is used for organizing rescue teams, material and food supply for trapped inhabitants. In this case a manned helicopter where the human pilot is reporting what the catastrophic area looks like to the expert system. Here it would be of advantage for the pilot to use natural language.

We describe two communication processes, one for an autonomous observer describing a constellation of objects, seen from the birds-eye perspective, in natural language to a human listener. The other for a person to describe an object constellation to a listening machine.

We want a description that is easily understood and constructed by a person and contains all necessary details but no further information. To reach that we feed the description produced by the autonomous observer to the autonomous listener to reconstruct the scene. This serves for us as a proof of concept, that this amount of qualitative detail in the described order is sufficient for either way of communication.

It might be of advantage if the listener, regardless of if he is a person or a machine, could direct questions to the observer to demand clarifying information. The advantage would be that the desired information can flow in at a point when the listener needs it. In this way the listener can continue his work without interruption to the already begun process.

Thus two communication models are thinkable for our approach. In the first case, shown in figure 1, we do not allow the listener to ask questions. The observer describes the scene in a certain order and the listener gets the description in the same order. Therefore care has to be taken to produce a coherent description to make the listener's work easier.

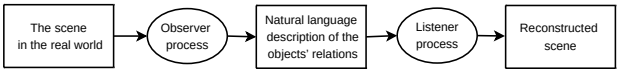

Figure 1. One-way communication model between the observer and the listener.

In the second possibility we allow the listener to ask clarifying questions during the ongoing process. The questions will interrupt the observer's description, and will be answered at once. We use this model that is illustrated in figure 2 in the remainder of the text within our examples.

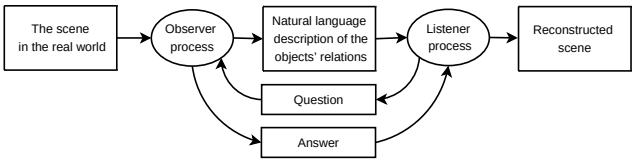

Figure 2. Two-way communication model between the observer and the listener. The listener can address questions directly to the observer, the observer answers directly to the listener.

5 Relational object position description

Our design is for a system that produces binary spatial relations between the objects it observes and communicates them to the listener which on the other side takes them as input and reconstructs a two-dimensional model expressing the relations between all the objects. The relations used are axis-parallel in a two dimensional coordinate system (in front of, to the left of, etc.) as well as diagonal relations in the same system.

To describe the objects in relation to each other we need a frame of reference [9], [7]. Therefore for each relation description we chose one of the present objects as reference object. All relations that are used in the system are defined in such a frame of reference around a reference object.

We call the model that we use and that is presented in figure 3 a double cross grid as it can be seen as a two-dimensional extension of Freksa's double cross calculus [4].

Freksa describes the double cross calculus for oriented objects. He uses a neighborhood-oriented representation to reason about spatial direction information. His approach deals only with point objects

whereas we need to model two-dimensional objects like cars seen from the birds-eye perspective.

In our case the reference object is in the middle of the grid and the plane around it is divided into nine areas. You can think of the grid's lines as extended lines of the bounding box of the reference object which is the same approach as in Mukerjee and Joe [8].

Mukerjee and Joe use a two-dimensional model for calculating position relations of objects. They use an enclosing box around the object that first has to have an identified front. They then divide the plane around this box by extending the boxes lines in eight two-dimensional regions that are named 1 to 8.

In our approach we deal with objects with orientation and we assume the objects' intrinsic front to the side where their orientation points to. According to the intrinsic front the regions that occur around the objects are named: straight front, right front, right neutral, right back, straight back, left back, left neutral, left front and identical. Identical indicates the same position as where the reference object is situated.

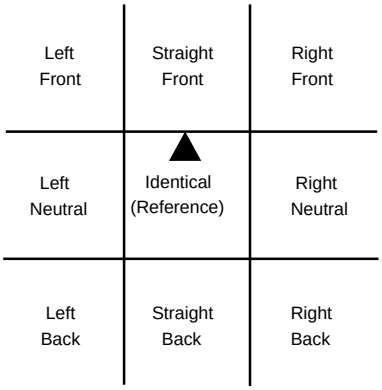

Figure 3. The plane around the reference object is divided into the nine regions straight front, right front, right neutral, right back, straight back, left back, left neutral, left front and identical. Another object's position can be described by giving the name of the of the reference object's qualitative region where the object to describe is in.

The position of another object would just be the name of the region it is in. It does not matter where it is within this region. As all regions except the region called identical are infinite the object can as well be arbitrarily far away from the reference object.

If an object is much bigger than the reference object and occupies several qualitative regions e.g. a truck beside a small car the appropriate relative position description would be all of those regions where the object is at least partly inside. A truck in relation to a car as reference object might be in the regions left back, left neutral and left front. On the other hand the car in relation to the truck as reference object would be just in the region right neutral.

When an object is not bigger than the reference object but still is partly within two qualitative regions we would state both the regions as its position description. The model used is further described in [13] and [14] where it is used to describe traffic maneuvers as chains of the qualitative states. Steinhauer [15] uses the same model as a basis for describing traffic maneuvers on different mental abstraction levels.

Several further approaches exist that use similar types of models for qualitative reasoning about directed and/or moving objects.

The direction relation matrix used by Goyal and Egenhofer [5] to calculate distances in similarity between spatial scenes forms a grid around the target object that divides the plane into nine quadrants where the eight surrounding quadrants are named north, northeast, east, southeast, south, southwest, west, and northwest.

One approach for tracking traffic, including learning traffic maneuvers from the observed information, from a video input taken by a stationary camera has been done by Fernyhough et al. [3]. The relative positions of objects that are close to the reference object are given as well as a grid around the reference object. The areas around it are named: Ahead, Ahead Right, Right, Behind Right, Behind, Behind Left, Left, and Ahead Left, which suited the traffic domain very well.

Distances between the objects are not taken into account in our model. In many applications though the distance between the objects might be crucial and determining if something in the scene is regarded as interesting or not. Distance can easily be attached to the model e.g. by giving a quantitative radius around each object and just be concerned with the other objects that are inside this radius. This might especially be of interest when dealing with moving objects.

Anyway we are not concerned with distance information at this moment as we just want to describe and reconstruct a scene of static or moving objects.

However, some implicit qualitative distance information is available from our model. It is part of the plan for or future work to make this distance information usable for applications.

6 Describing an object constellation

To explain how the observer describes the scene we will consider the object constellation shown in figure 4a) Suppose the observer discovers and identifies the objects in the order of their numbering 1, 2, 3, 4, and 5. In reality an observer would usually come up with a more natural ordering of the objects. If he would scan the territory in an intuitively more logically planned way he might give the order 3, 5, 4, 2, and 1 or 4, 1, 5, 3, and 2.

The order in which the observer discovers the objects influences the order in which he communicates their relations. The last two given orderings would make the reconstruction of the scene very easy as new discovered objects would just need to be added at the outside of the already reconstructed scene. This is in fact an advantage as the reconstruction effort is kept very low. In the following example we would however like to show what the system is capable of doing and therefore choose a more unstructured order of objects that will force the system to retract previously made decisions.

The observer produces the objects' relations in the order we mentioned above. The first information states that he found an object which is the first object (1). As no other objects are found yet no further information can be given. Observe that the numbers the observer assigns to the objects are chosen in the order in which he discovers the objects and has nothing to do with the objects "names". In this example the objects in the original constellation in figure 4a) are numbered already in the order they are observed.

When object 2 is detected its relation to the object detected before will be calculated and the relation is given as (2 right neutral 1) which means that object 2 is right neutral of object 1. As the objects in this example do not have any orientation the observer will attach the orientation of north or up to all of them. This is necessary to orientate the double cross grid around the objects. When object 3 is discovered the relations (3 right front 1) and (3 left front 2) are added. The detection of object 4 adds the further relations (4 straight front 1), (4 left front 2) and (4 left back 3). For object 5 the additional relations (5 right front 1), (5 left front 2), (5 left back 3) and (5 right front 4) are

added. These eleven relations are sufficient to reconstruct the scene with all qualitative relations between the objects preserved.

7 Reconstructing an object constellation

The constellation of objects is considered as an object allocation on a two-dimensional grid. It would be possible to place the objects arbitrarily far away from each other as long as no relations that limit the distance between the objects are included. Thus there exists an infinite number of configurations that are consistent with any given set of input relations.

The basic idea in our approach is to impose a preference on the possible configurations so as to prefer configurations where objects are adjacent to each other, or as close to each other as possible. We motivate this default on two grounds: the grid is anyway purely qualitative and no metric is assumed on it, and the default appears to result in reasonable answers to the queries that can be put to the system. The default is applied incrementally. The first received relation constructs a configuration that is consistent with the default. Further relations that are added to the system accumulate information to the configuration but may even cause the system to retract previously made default decisions in order to obtain a configuration that conforms to the given input.

Incorporating an additional relation into a configuration is done as follows. First the qualitative region of the given reference object where the new object needs to be placed has to be located. After that the relations of all objects that are in a horizontal or vertical in-line region are regarded.

In-line regions are those qualitative regions of the same reference object that are horizontally or vertically in a line with the region where the new object has to be placed. If the new object is for example right front of the reference object the regions straight front and left front are horizontally in-line with that region and the regions right neutral and right back are vertically in-line with it.

Objects that are placed within those regions might influence the position of the new object to place. We will call these objects the influencing objects. It needs to be checked for each of them how the relation between that object and the new object will be.

According to those relations the possible space for the new object might be narrowed down further and further and with it the number of remaining influencing objects will often decrease as well. The number of relations to consider can be further reduced by the order in which the influencing objects are checked against the new objects position.

Sometimes some space has to be made between some already placed objects e.g. when the new object happens to be in the middle of them (e.g. left of A and right of B where B already is left of A).

Of course care has to be taken when objects are moved that already existing relations are not accidentally changed to wrong. Therefore the objects in the picture are divided into two groups, one group containing all the objects that will be on the one side of the new object and the other group containing all the objects that will be on the other side of it. With this done the object groups can be moved further apart from each other without changing any relation of any pair of objects.

8 A simple example

The example in figure 4 illustrates the idea. To keep the example simple the objects have no given orientation and we attach the same intrinsic front to them which in this case means that they are all facing north. Figure 4a) shows the object constellation that has to be communicated and reconstructed.

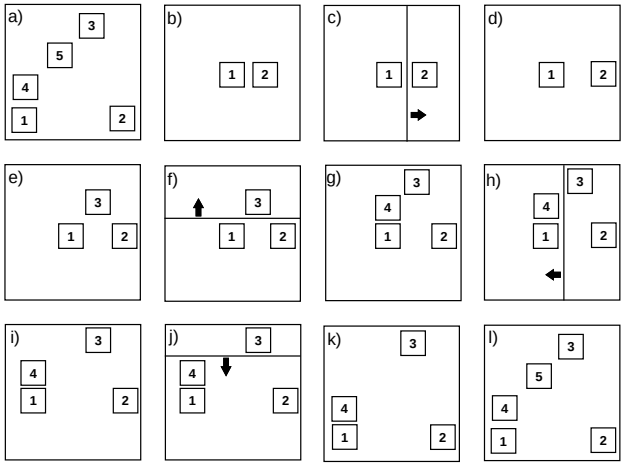

Figure 4. a) The original constellation that has to be reconstructed b) Object 1 is placed somewhere (it does not matter where because we are only interested in the qualitative position relations between the objects). Object 2 is placed right neutral of object 1. c) Object 3 has to be inserted between the objects 1 and 2 therefore object 2 is moved outwards to the right. d) Free space appeared after moving object 2 outwards. e) Object 3 can now be inserted at the right position. f) Space has to be arranged for object 4 that needs to be above the objects 1 and 2 but below object 3. g) Object 4 is now placed at the right position. h) Object 5 needs to be right of the objects 1 and 4 therefore these objects are moved outwards to the left. i) Free space appeared to the right of the objects 4 and 1. j) Object 5 also needs to be above object 4 and below object 3. Therefore all objects that are on the lower side of the line between 4 and 3 are moved downwards to make some space. k) The space is arranged for object 5. l) Object 5 is placed and the original constellation is thereby reconstructed. All relative position relations of the objects are the same as in the original in a).

The listener gets the stream of relations in the order the observer produces it. We use the description of the scene in figure 4a) whose construction was already described in section 5.

The input (1) causes the listener to assume that there is an object and he will place an object somewhere in his picture. The next relation he learns is that object 2 is right neutral of object 1. Object 2 can be placed directly and it will be placed very close to the reference object (figure 4b)). The next relation is object 3 right front of object 1. If object 3 is in the right front region of object 1, object 2 is in a vertically in-line region. (Object 2 is right neutral of object 1.) Therefore the listener must know the relation between object 3 and object 2 and asks the observer to provide the actual value of this relation. The observer answers (3 left front 2). That means that object 1 will be on the left side of object 3, and object 2 will be on the right side of object 3. In figure 4c) we drew a line to indicate where the objects have to be separated. All objects that are on the right side of this line will be moved outwards to make some space. The result is shown in figure 4d). Free space appeared where object 3 can be inserted, which is done in figure 4e). The next object to place is object 4, which is straight front of object 1. Being there, object 3 is in a region that is horizontally in-line and we need to know that object 4 is left back of object 3. That means that object 4 will end up being somewhere in front of object 1 but back of object 3. Therefore we draw a line between object 3 and 1 and move everything above this line upwards (figure 4f)). Free space appears where object 4 can be placed and all other relations of the other objects in the picture stay the same (fig-

ure 4g)). Now object 5 needs to be placed right front of object 1. It will be influenced by the objects 2, 4 and 3. That means we need to know that object 5 is right front of object 4, left back of object 3, and left front of object 2. From this we get that object 5 has to be merged in to the left of the objects 3 and 2 and to the right of the objects 1 and 4. Therefore we draw a vertical line and move the objects left of it further to the left to obtain some new space (figure 4h) and 4i)). In addition we need to draw a horizontal line beneath object 3 and move all the objects under this line downwards (figure 4j)). Figure 4k) shows the resulting space after we have moved the objects, finally in figure 4l) object 5 is placed and we have reconstructed the scene in figure 4a).

9 Objects with orientation

The same process can be used for a constellation where the objects have different orientations. However in that case, we would adopt the normal way of speaking and say that an object that is east of another object that itself is facing north is to the right of the object. If the second object would be facing south the object being east of it would be to the left of the second object. That means the the reference object sets the frame of reference that is used for the relation description. As the reference object has an intrinsic front this is used for the frame of reference. If we change from one reference object to another the frame of reference changes with it.

If we add orientation information to our example above but do not change anything else we obtain the structure shown in figure 5.

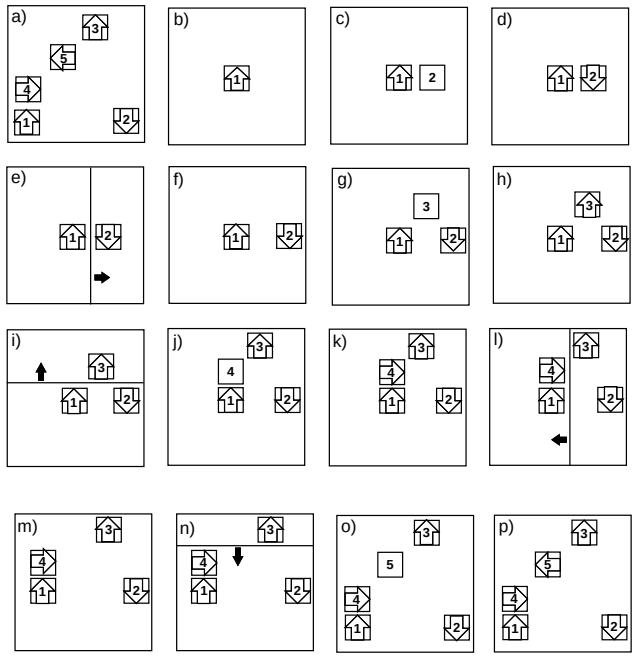

Figure 5. a) The object constellation to reconstruct. b) Object 1 is placed in the scene with an arbitrary orientation. c) Object 2 is placed right neutral of object 1 and d) object 2's orientation is added. e) and f) show the process of making some space to add object 3 right front of 1 and right back of 2. Object 3 is added in g) and its orientation in h). The pictures i), j), and k) show the process of adding object 4 and the rest l) to p) the process of adding object 5.

The observer detects object 1 first and states that there is an object (1). The orientation of the object so far does not matter but when object 2 is discovered we need the relative orientation information. The observer would state that object 2 is right neutral of object 1 (2 right neutral 1) and that would set an implicit relative orientation to object 1.

Note here that the orientation could as well be given in a global frame of reference like north, south, east, west or up, down, left, right if that would be necessary in any way for the application and if the information would be available. In this approach we would however like to ignore the global frame of reference and just concentrate on what we can gain from the implicit information from the objects at hand.

Thus in our example we still do not know which orientation object 2 has. Therefore one further relation has to be given, namely the relation of one object to object 2 that we already know about and that we already gave an implicit orientation. In this case the only such object available is object 1 and the observer is forced to say that object 1 is right neutral of object 2 (1 right neutral 2).

The discovery of object 3 will lead to (3 right front 1) (3 right back 2) and (1 left back 3) where the last relation gives information about object 3's orientation. Object 4 is added by the relations (4 straight front 1) (4 right back 2) (4 left back 3) and for the orientation information of object 4 (1 right neutral 4). At last object 5 contributes with the relations (5 right front 1), (5 right back 2), (5 left back 3), (5 left front 4) and (1 left front 5) for the orientation information of object 5.

The listener gets the information in the same order as the observer produced them and reconstructs the scene step by step. First he places object 1 into the scene (figure 5b)). He is free to choose an orientation for object 1. As we are only concerned about relative positions the orientation that will be attached to object 1 does not influence the result. Regardless of which orientation the listener chooses, all relations in the scene will in the end be the same as in the original from figure 3a), but the scene might be turned as a whole.

To keep the example simple and make it easier to compare the reconstructed scene in every step with the original scene we suppose that the listener chooses the same orientation as in the original. If there would be the need to develop the scene in an overall global frame of reference the observer only needs to pass along the orientation information in the global frame for the first object. In this case the orientation would be north or up, depending on what global frame of reference is used.

Secondly object 2 is right neutral of object 1 and can be placed there (figure 5c)), the third relation clarifies the orientation of object 2. If object 1 is right neutral of object 2 at the same time as object 2 is right neutral of object 1 then the orientation of object 2 must be the opposite of object 1 (figure 5d)).

The process goes on in the same way as before. Object 3 has to be placed right front of object 1 and right back of object 2. Unfortunately there is no space available so far, as the default was kept to place new objects in the constellation as close as possible to the old objects. Now the scene is split vertically between the objects 1 and 2 (figure 5e)), where new space is needed and the two resulting groups of objects are moved apart from each other to make some space for object 3 (figure 5f)).

In figure 5g) object 3 is placed and figure 5h) gives the result of the evaluation of (1 left back 3) which clarifies object 3's orientation.

The rest of the process is the same as before, the region for the new object is calculated, space is made if necessary by dividing the objects into two groups horizontally or vertically between the objects where the space is needed, and the new object is placed. After that the orientation of the new object is added.

We illustrated how an object constellation can be described and reconstructed. In our examples the relative position that is named first to introduce a new object is always given to the first object. That is not necessary. Any object that has already been mentioned before can be used as a reference object. The intermediate results during the scene development are then of course different from our example but the result in the end will be qualitatively the same.

At any time during the process the objects' relations in the reconstructed picture will be the same as in the original scene.

The orientation information that is always given last in our example, when the final position of the object is found, can be given earlier and does not have to be given in relation to the first object. Any object already in the scene can be used as reference object to give the orientation information for the new object.

10 Moving objects

So far we were only concerned with static objects with or without orientation. Now we expand our approach to moving objects.

When we observe moving objects, the relative positions between them will change over time. The observer has the duty to keep track of the changes and notify them to the listener so that he can update his reconstructed image whenever relevant changes have occurred.

We assume that the observer uses a tracker to keep track of the once identified objects. Every single object trajectory contains a lot of quantitative data but not all of it is relevant in terms of communication. As we stated before, the listener just needs to know the objects' relative positions. Once he has been given those and as long as they do not change there is nothing further to communicate.

Only when the objects' qualitative relations change the change has to be announced. This will reduce the amount of data exchange that the listener has to deal with radically. In figure 6 two objects are shown together with their double cross around them. Whenever an object passes over one of the lines in the picture a change of the qualitative relation between the two objects has occurred and only then will the observer announce an update on the scene.

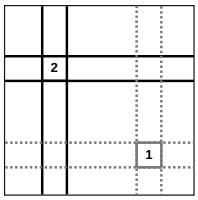

Figure 6. Two objects shown with their double cross grids. No qualitative change will take place as long as the objects stay in the other object's region they are now. Only when an object crosses a line belonging to the other object's regions a qualitative change has occurred.

Whenever a qualitative position change occurs this will be announced as soon as possible by the observer. The listener assumes that nothing else than the announced information has changed and will update the reconstructed picture step by step in the same order as the update information arrives.

There is the possibility though that two or more changes are caused by the same object and therefore occur at the same time (see figure 7a)). When object 3 moves to the left it will change the relative position to object 2 from straight back to left back and at the same time change its relative position to object 1 from right back to straight back.

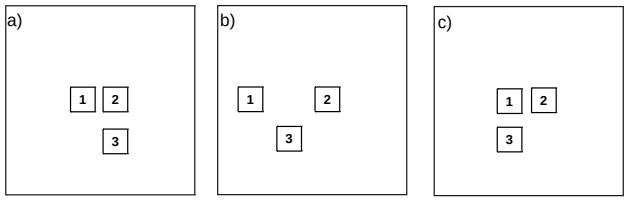

Figure 7. a) Object 3 is straight back of object 2 and left back of object 1. b) the situation after the listener got the information that object 3 is now left back of object 2 but did not get any further information that causes him to change the relation of object 3 to object 1 as well. c) The updated situation if the further information that object 3 is now straight back of object 1 is found together with the information that object 3 is now left back of object 2.

The observer has no other choice than stating first one change and then the other in linear order. The listener will however realize that the change of object 3 from straight back of object 2 to left back of object 2 could also mean that a change of the relative position of object 3 to object 1 might have taken place and will ask for clarifying information. If no such change is given the listener will organize the reconstructed picture according to all valid relations which will result in the scene shown in figure 7b).

But when additional information is given that states that object 3 is straight back of object 1 as shown in figure 7c) the listener needs just to place object 3 at its new position without rearranging the rest of the scene. He already took care of some further update that he would have to do anyway and he saved the time to construct a temporary and furthermore wrong picture.

When many objects change their relative positions fast the scene needs to be updated constantly. The observer will give the updated information as soon as he realizes a qualitative change in the scene. Ideally the changes will be given in the exact order in which they are realized, which has the consequence that the reconstructed scene on the listener's side will develop in the same way and order as the original scene. Changes that occur at the same time must however be linearized though in order to communicate them but they will be mentioned as closely as possible.

11 An example with moving objects

In the following example in figure 6 we have three objects with orientations in the scene. We assume that the observer will first describe the scene from figure 8a) and then update the scene from time to time.

The listener will reconstruct the scene step by step and after that continue to change the scene according to the update information he gets from the observer.

11.1 At the observer's side

The first output from the observer would be (1) to name the first object, (2 right front 1) to give the information where object 2 is in relation to object 1, (1 right back 2) to clarify which orientation object 2 has, (3 right front 1) and (3 left front 2) to give the information where in the scene object 3 fits in, and (1 left front 3) to clarify object 3's orientation.

We will now explain how the scene will be updated according to the objects' movement. We therefore assume that the objects move in the direction given by their orientation and all with about the same speed. Figure 8b) shows the scene after a while. Until now no qualitative relation between the objects has changed and therefore no update has to be given.

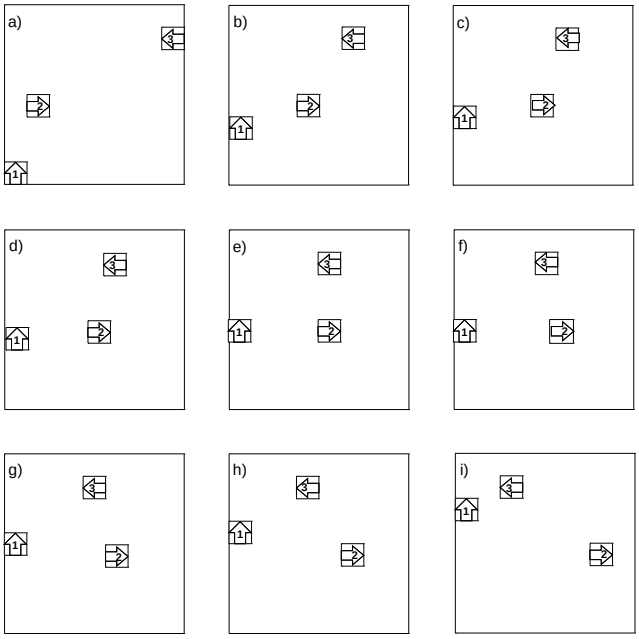

Figure 8. a) The scene in the beginning of the observation b) The objects have moved but no qualitative change has occurred therefore no update has to be given. c) Update1: (1 right back & straight back 2) d) Update2: (2 left neutral & left front 3) e) Update3: (1 straight back 2), (2 left neutral 3) f) Update4: (1 straight back & left back 2), (2 left neutral & left back 3) g) Update5: (2 left back 3) h) Update6: (1 left back 2) i) The objects have moved further but no qualitative changes have occurred.

A little later the scene looks like in figure 8c) and now one change has occurred when object 1 changed its qualitative position from right back of object 2 over one of the double cross lines of object 2, and is now partly in the straight back region and partly still in the right back region of object 2. The observer will state this fact as (1 straight back & right back 2). All other qualitative relations are still the same as in the first description.

The next change to announce is when the objects 2 and 3 begin to become parallel to each other (figure 8d). Object 2 has begun to enter object 3's left neutral region and object 3 has of course at the same time begun to enter object 2's left neutral region. The change to state is (2 left front & left neutral 3). This is enough information as it implicitly contains the information (3 left neutral & left front 2) which itself would be sufficient to state alone.

As the objects move on two more changes take place at the same time. Object 1 is now straight back of object 2 and object 2 is left neutral of object 3 which of course again implies that object 3 is left neutral of object 2. Therefore only two changes have to be announced to update to the scene shown in figure 8e) which are (1 straight back 2) and (2 left neutral 3).

After a while, in figure 8f) the objects begin to leave those regions again and the result will be expressed as (1 straight back & left back 2) and (2 left neutral & left back 3).

In figure 8g) the next change has taken place as object 2 and 3 have now completely passed each other (2 left back 3) is enough information to explain this change. Though object 1 has moved as well its qualitative relation to the other objects has not changed.

In figure 8h) you see that object 1 has passed object 2 and is now left back of it. Still its qualitative relation to object 3 has not changed and (1 left back 2) is the complete update information for this time.

If all the objects move on in the same way there will be no further changes until the state shown in figure 8i) is reached.

At this time it should have been clear to the reader when and in what way the update information is found and formulated and we conclude the example at this point.

11.2 At the listener's side

The listener on the other side will take one information at the time and weave it into the scene. Thereby the scene will be updated in the same way as the observer states the changes.

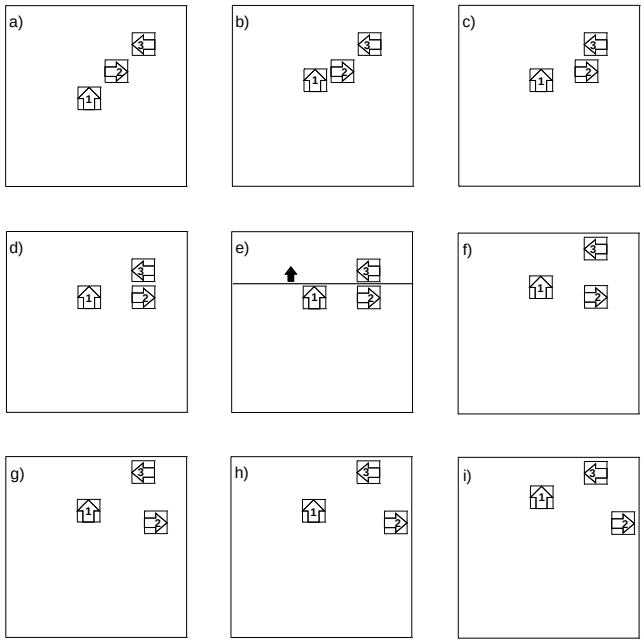

Figure 9. a) The scene after the first description of the constellation b) Update1: (1 right back & straight back 2) c) Update2: (2 left neutral & left front 3) d) Update3: (1 straight back 2), (2 left neutral 3). e) The first entry in Update4: (1 straight back & left back 2), (2 left neutral & left back 3) causes the listener to make same place to correctly place object 1 at its new position. f) The first entry in Update4 has been taken care of. g) Even the second entry of Update4 has been progressed completely. h)Update5: (2 left back 3) i) Update6: (1 left back 2)

The scene description Scene: (1), (2 right front 1), (1 right back 2), (3 right front 1), (3 left front 2), (1 left front 3) will course the listener to establish a scene in the same way as described before. He places object 1 somewhere in the scene, then attaches an orientation to it, as no global orientation is given which could be applied. Object 2 is placed as close as possible right front of object 1 and the orientation that is conform with the fact that object 1 is right back of object 2 is added. After that object 3 needs to be placed in the right front region of object 1 which has to be narrowed down due to the influencing object 2 that already is in the same region. Object 3 is placed and its orientation, conform with the information that object 1 must be left front of object 3 is set. The result of this procedure is shown in figure 9a).

The first update Update1: (1 straight back & right back 2) means to the listener to leave everything as it is except for the position of object 1 which is now straight back and right back of object 2 which is shown in figure 9b).

The seconde update shown in figure 9c) Update2: (2 left front & left neutral 3) also courses just one change. Object two is moved a

little and all other objects positions and their relations are left the same as before.

After the third update Update3: (1 straight back 2), (2 left neutral 3) the scene on the listener's side looks like in figure 9d) where the objects 2 and 3 are parallel and object 1 is now completely straight back of object 2.

By analyzing the first entry in Update4: (1 straight back & left back 2), (2 left neutral & left back 3) the listener realizes that in his reconstructed picture the change of object 1's position from straight back of object 2 to straight back and left back of object 2 would imply that the relative position of object 1 to object 3 as well has been changed. The listener asks the observer about the relation of object 1 to object 3. The answer is (1 left front 3) which clarifies the situation. A change of the relation was not confirmed and therefore the listener needs to change the reconstructed picture in the way that the new relation (1 straight back & left back 2) as well as the old relation (1 left front 3) can coexist. Therefore some space has to be made which is indicated in figure 9e). This follows the same procedure as in the processes described before whenever an object had to be placed somewhere where no space was available so far. Figure 9f) shows the situation after the first entry in Update4 has been worked in. In figure 9g) even the second entry of Update4 has been handled.

Update5: (2 left back 3) will result in the picture in figure 9h) as will Update6: (1 left back 2) cause the listener to construct the new scene shown in figure 9i).

12 Conclusions and discussion

We provided a method with that an autonomous observer of a scene of static objects can qualitatively describe the object constellation. The method works for objects with and even without intrinsic orientations.

Further on the same method was applied to describe and update a scene of moving objects. The resulting qualitative description is easily understandable for a human user and is sufficient as a qualitative relation description of the objects.

Next we described how the same approach can be used for an autonomous listener to reconstruct the scene from the given description.

In our example the order in which the observer picked the objects was chaotic. A more intuitive order would reduce the amount of adjustments needed during the reconstruction process.

The methods presented here will be of advantage whenever humans and machines need to communicate spontaneously on the basis of human understanding of how to describe an object scene in natural language.

However the question arises how much communication between the observer and the listener is desirable. The two extreme cases would be a) that the listener is not allowed to ask at all and has to deal just with the information the observer provides. A process the listener has begun might have to be postponed according to, at the moment, missing information.

The other extreme b) that the observer only answers to questions is not a good idea, as the listener does not know what to expect in the scene and would have difficulties to ask after all the things he does not even know exist.

However, communication seminars for people show that reconstruction of a scene that one of the persons sees and the other have to reconstruct from verbal description, wins enormous if the listeners are allowed to ask questions.

The reconstructed picture at the listener's side will be qualitatively correct regarding the so far transmitted relations, but defaultly placed objects might have to be moved if new information comes in. Clarifying information might be needed to place a new object correctly into the existing picture or to retract earlier made decisions.

If we would use machine-machine communication, we could use a hearer model inside the observer to keep track of what the listener knows and we could use the same program for the listener and the hearer model.

In our case though one of the participants is a human being, and even if he is fully capable of following a certain procedure of describing the objects or reconstructing the scene a certain amount of freedom, to take decisions in his own way of thinking, seems to be appropriate.

The examples show that the qualitative information that can be gained from the double cross method is sufficient for describing, reconstructing and updating a scene of moving objects. However this does not mean that the results can be applied in arbitrary applications. Our results are purely qualitative and no quantitative information about speed, distance etc. is available.

13 Future Work

So far only objects of the same shape and same size have been considered. However qualitative size information is already implicitly available from the process.

Imagine three objects that are in a line. That means object 2 is right neutral of object 1 and object 3 is right neutral of object 2. So far there are no other objects to consider so that the three objects are as close as possible to each other. Now object 4 has to be merged into the scene and the observer tells the listener that object 4 is straight back and left back of object 3 and also left back, straight back and right back of object 2 and as well straight back and right back and left back of object 1. This means that object 4 must be larger than any of the three previously mentioned objects.

As well qualitative distance information of the objects is already implicit in the model. The number of objects that are in between to objects can give an idea of how far the objects are away from each other. Imagine three objects, where object 2 is right back of object 1 and object 3 is right back of object 2. It is clear that object 3 must be further away from object 1 than object 2.

In the examples so far the observer had the full overview of the scene at all time points and was able to announce a change as soon as it took place. In the future we are going to model scenes where the observer has a field of view that is smaller than the whole scene. Updates are given, when the observer detects a change. Consequences of this are that some qualitative relations between the objects can be missed. However, the conceptual neighborhood of qualitative states will make it possible to conclude which relations the objects must have had in between.

During the reconstruction of the scene objects that are in horizontally or vertically in-line regions have an influence on the position where a new object has to be placed. With each influencing object that we check against, the space for the new object to place is narrowed down and with it most likely the amount of remaining influencing objects. A wise choice of the order in which the influencing objects are taken into account can reduce the amount of influencing objects drastically. We will investigate the question whether this wise choice can be automated.

REFERENCES

[1] M. Arens and H.-H. Nagel, 'Behavioral knowledge representation for the understanding and creation of video sequences', in *Proceedings*

of KI 2003: Advances in Artificial Intelligence, eds., Andreas Günter, Rudolf Kruse, and Bernd Neumann, LNCS, pp. 149–163, Hamburg, Germany, (September 15-18 2003). Springer.

[2] P. Doherty, G. Granlund, K. Kuchcinski, E. Sandewall, K. Nordberg, E. Skarman, and J. Wiklund, 'The witas unmanned areal vehicle project', in *Proceedings of the 14th European Conference on Artificial Intelligence*, pp. 747–755, Berlin, (August 2000).

[3] J. Fernyhough, A.g. Cohn, and D.C. Hogg, 'Constructing qualitative event models automatically from video input', in *Image and Vision Computing*, volume 18, pp. 81–103, (2000).

[4] C. Freksa, 'Using orientation information for qualitative spatial reasoning', in *Proceedings of the International Conference GIS-From Space to Territory: Theories and Methods of Spatio-Temporal Reasoning*, LNCS, Pisa, (September 21-23 1992).

[5] R. K. Goyal and M. J. Egenhofer, 'Similarity of cardinal directions', in *SSTD 2001*, ed., C. S. Jensen, LNCS 2121, pp. 36–55. Springer-Verlag Berlin Heidelberg, (2001).

[6] G. Herzog and P. Wazinski, 'Visual translator: Linking perceptions and natural language descriptions', in *Artificial Intelligence Review, 8*, pp. 175–187, (1994).

[7] S. Levinson, 'Frames of references and molynexs question: crosslinguistic evidence', in *Language and Space*, eds., P. Bloom, M. Peterson, L. Nadel, and M. Garret, pp. 109–169. MIT Press., (1996).

[8] A. Mukerjee and G. Joe, 'A qualitative model for space', in *Proceedings of the AAAI*, pp. 721–727, Boston, (1990).

[9] G. Retz-Schmidt, 'Various views on spatial prepositions', in *AI Magazine*, volume 9, pp. 95–105, (1988).

[10] E. Sandewall, P. Doherty, O. Lemon, and S. Peters, 'Words at the right time: Real-time dialogues with the witas unmanned aerial vehicle', in *Proceedings of the 26th German Conference on Artificial Intelligence*, eds., R. Kruse A. Gunter and B. Neumann, volume 2821 of *LNAI*, pp. 52–63, (2003).

[11] Erik Sandewall, Patrick Doherty, Oliver Lemon, and Stanley Peters, 'Real-time dialogues with the witas unmanned aerial vehicle', in *Annual German Conference on AI*, ed., Andreas Günter, pp. 52–63, (2003).

[12] R. Sproat, 'Inferring the environment in a text-to-scene conversion system', in *Proceedings of the First International Conference on Knowledge Capture (K-CAP 2001)*, pp. 147–154, Victoria, BC, Canada, (October 21-23 2001). ACM.

[13] H. J. Steinhauer, 'The qualitative description of traffic maneuvers', in *Proceedings of Workshop on Spatial and Temporal Reasoning*, eds., H. W. Güsgen, F. D. Anger, C. Freksa, G. Ligozat, and R. V. Rodrguez, 16th European Conference on Artificial Intelligence (ECAI), pp. 141–148, Valencia, Spain, (August 23-27 2004).

[14] H. J. Steinhauer, 'A qualitative description of traffic maneuvers', in *Spatial Cognition 2004: Poster Presentations*, eds., T. Barkowsky, C. Freksa, M. Knauff, B. Krieg-Brückner, and B. Nebel, Report Series of the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition, pp. 39–42, Frauenchiemsee, Germany, (October 2004). SFB/TR 8 Spatial Cognition.

[15] H. J. Steinhauer, 'A qualitative model for natural language communication about vehicle traffic', in *Reasoning with Mental and External Diagrams: Computational Modeling and Spatial Assistance*, ed., M. Hegarty R. Lowe T. Barkowsky, C. Freksa, pp. 52–57, Stanford, California, (March 21-23 2005). AAAI Press.

[16] H. J. Steinhauer, 'Towards a qualitative model for natural language communication about vehicle traffic', in *IJCAI-05 Workshop on Spatial and Temporal Reasoning*, ed., G. Ligozat H. W. Güsgen, C. Freksa, pp. 45–51, Edinburgh, Scotland, (July 30th - August 5th 2005).