Towards a Qualitative Model for Natural Language Communication about Vehicle Traffic

H. Joe Steinhauer
Department of Computer Science
Linköping University
S-58 183 Linköping, Sweden
joest@ida.liu.se

Abstract
In this paper we describe a technique how an object constellation can be successively reconstructed just from its qualitative position relation descriptions. A simple and intuitive model functions as a frame of reference for both, the relation descriptions and the reconstruction of the object constellations.

1. Motivation

Imagine a group of agents that work together in a disaster area. The group consists of human experts that are needed in the specific type of catastrophe, e.g. rescue after an earthquake or fire specialists in a quick spreading forest fire, together with autonomous robots like all-terrain vehicles and helicopters that explore the area and report where what kind of help is needed and also distribute needed tools to rescue parties on the field or survival kits to trapped inhabitants.

The communication between the different kinds of agents takes place in natural language. This is advantageous because the human experts are only experienced in their specific rescue field; they are not required to have further training how to communicate with the autonomous agents. Thus the experts can concentrate just on their work instead of the communication; and every available expert can be deployed at the catastrophic field and not just further communication trained experts. Therefore the problem how the communication is done has to be solved on the autonomous agents' side; those have to learn to communicate with the experts, using everyday language with the terminology of the rescue field, and not the other way around.

In the continuation of this paper we concentrate on the case that an observing agent wants to communicate how the scene it observes looks like. That means to communicate a constellation of objects without using quantitative information, like x-y-coordinates in a metric system or distances. The listener shall be able to reconstruct the scene where the reconstruction matches all the qualitative aspects of the original scene. We claim that qualitative relation information is enough to get a matching picture. At first we will point out which is the relevant qualitative information that people would use to describe such a scene, then we introduce a simple and intuitive model that suffices just these purposes and further on we show a technique how an object constellation can be reconstructed only using their relative position information.

2. Qualitative aspects of natural language communication

Naturally people communicate in natural language. Even though it is ambiguous and from time to time misunderstandings do occur it still is the best general-purpose communication tool we can think of. We humans are highly skilled in describing and explaining complex situations and relations in natural language. As we do not have any other natural way, except for maybe drawing, to send on our sensor data, we need to be. To produce a drawing we need further tools and we need to pass the drawing on to the other person if he is not at the same location. Thus even though we sometimes feel that a picture is worth more than a thousand words, we do not always have a picture available and then we need to rely on language again.

If we want to describe a visual scene e.g. a constellation of objects we will not describe every pixel of the picture. Instead we group, classify and abstract information. We recognize objects like cars or houses; we group them together to queues of cars or blocks of houses and we explain their positions in relation to each other. Furthermore we classify them as small or big or smaller or bigger than some other object. We express distance in terms of nearness or farness and speed in terms of everything between very slow and very fast without mentioning the actual speed of the object. That means that we mainly use qualitative information for the description.

In other words we first recognize objects in the picture; we then abstract what we saw in a certain way and keep this abstraction, the mental image, of the situation in mind. If we need to describe the situation to somebody else we mainly communicate this mental image.

Mental images are often discussed in the context of somebody remembering a visual input that is not visually available anymore. In our case we still have the visual input available but we need to communicate it to somebody who does not. We believe that the information picked to describe the scene to somebody else is about the same that would be chosen to keep in a mental image, namely just that information that the describer consciously or subconsciously chose to be important. The listener on the other hand needs now to rebuild the described picture in his mind by extracting the important information from the natural language sentences and build his own mental image of the situation. What the important information is can of course vary from person to person, from situation to situation and from task to task.

Rounding up, people extract the important, mainly qualitative information from a situation and organize them in a mental image, which abstract from all unnecessary details. This image is translated into a natural language description of the original situation. The description has to be decoded into a mental image of the listener and hopefully all the important information will turn out to be the same in the two mental images.

If we want a robot to communicate with a person in the same way the robot must first be able to use qualitative information. That means it needs to be able to abstract from quantitative details into qualitative classes in the same way people do and it must be able to understand qualitative information and to reason with it when it gets a natural language description of a situation. Thus we have to deal with qualitative reasoning in the first place.

> "Qualitative reasoning is the area of AI which creates representations for continuous aspects of the world, such as space, time, and quantity, which support reasoning with very little information." (Forbus, 1996).

Second the chosen qualitative information has to be described in natural language. In this case the robot has to be very flexible in understanding natural language descriptions as it shall be able to work with communicationally untrained human experts. Though of course the natural language vocabulary will be task specific, there still is a variety of different descriptions of the same situation. The choice lies by the human expert who chose it accordingly to his personal interpretation. The robot has to adapt to this interpretation and use the same vocabulary as the expert.

> "Natural-language and multimedia dialogue with an autonomous robot is a challenging research problem which introduces several important issues that are not present in, for example, dialogue with a database or a service provider such as an automated travel agency" (Sandewall et al., 2003, page 55).

Third, autonomous agents are often expected to draw some conclusions on their own. Therefore they need to do reasoning. Reasoning rules are often formulated in natural language. Experts often describe their expertise even in rules of thumb or default rules. Those rules correspond to specific abstraction levels that the experts use. Qualitative knowledge plays a major role in their decision processes as:

> "Qualitative knowledge can be viewed as the aspects of knowledge which critically influences decisions" (Freksa and Röhrig, 1993).

3. A qualitative model

To build a communication that is natural for the human experts the robots have to learn human communication. They have to use natural language to express the data they want to submit. Before that, they even have to extract the qualitative information that a human listener will understand from all their quantitative sensor input. There is no meaning in passing on the data stream the robot gets from its visual

sensors, a person will not be able to understand it anyway, especially not an untrained person that at the moment needs to fully concentrate on a different topic.
The approach we use is to first let the robot extract the qualitative information from its quantitative sensor input and second formulate it in a natural language description that it passes on to the human listener. The robot must also be able to understand a natural language description and extract from it the important information to get a picture of the situation described.
That means that we take the mental image of the situation as the basis for the communication. We want to achieve that the mental image of the robot is "similar" to the mental image of its human communication partner. An illustration hereof is given in figure 1. With similar we mean that the qualitative information that both use is the same in both mental abstractions. Of course we do not know how the human mental image looks like, and have to rely on the description that the human produces. Furthermore several different descriptions are possible to cover the qualitative information of one and the same situation but in the end all qualitative relations between the objects should be clearly captured in whatever description.

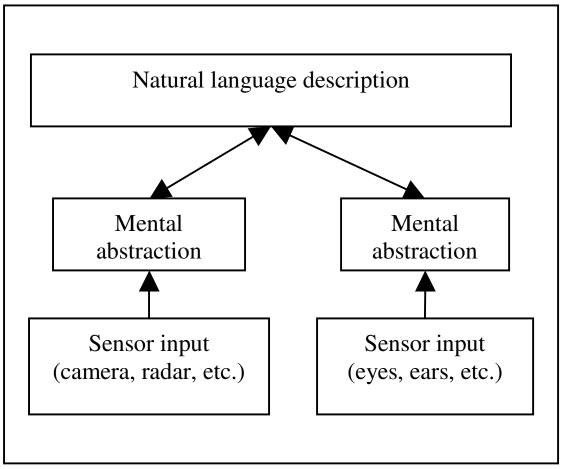

Figure 1 People and robots use quite different sensor input. From those they both have to extract a "mental abstraction" that captures the important information for the natural language communication.

4. Relational object position descriptions

When people describe an object constellation they describe the relations of the objects to each other. We use a very simple and intuitive model to extract the qualitative relations from a visual input. The model shown in figure 2 divides the plane around a reference object in nine separate qualitative regions including the region of the reference object itself. In this case the reference object has an intrinsic front so that the regions all can be identified by its names.
Reference objects are necessary for us to orientate our picture of the world. We say that they define a frame of reference (Retz-Schmidt, 1988; Levinson, 1996) according to which we then can describe the position of the object that we are interested in. If we know where the church is and which side of the church its intrinsic front is (the one with the big main entrance) we understand the sentence "The children are playing in front of the church". The church gives a frame of reference within the children can be placed.
In our model the position of another object has to be given according to the reference object. The position is the field where the object is placed in. To describe this relation we just have to give the name of that qualitative region. In the case of a moving object, as long as the object moves within the same region, nothing, qualitatively speaking, changes. In the case when it leaves one region and enters another an important (qualitative) change has occurred. When an object happened to be on a line that means that the

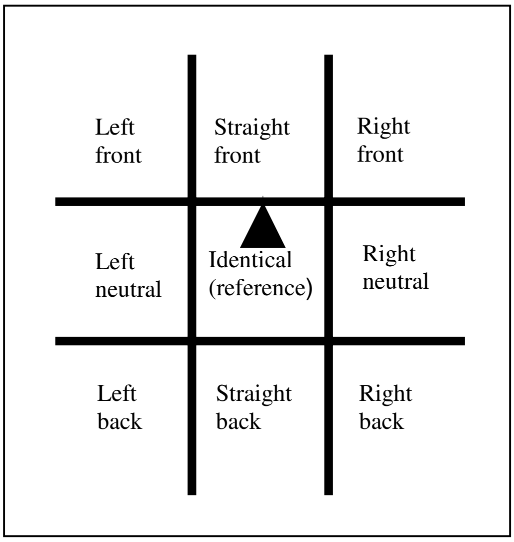

Figure 2 The plane around the reference object is divided into nine qualitative regions including the region where the reference object is itself. Another objects position can be described now by just giving the name of the qualitative region it is in.

object is actually in two regions at the same time we have to say that the object is in both those regions. The same is valid when an object is much bigger than the reference object and occupies more than one region at the time. The model is further described in Steinhauer (2004a and 2004b) where it is used to describe traffic maneuvers as chains of qualitative states. In Steinhauer (to appear 2005) the same model functions as basis for describing traffic maneuvers on different mental abstraction levels.
Several approaches exist that use similar types of models for qualitative reasoning about directed and/or moving objects. Freksa (1992) describes the double cross calculus for oriented objects. He uses a neighborhood-oriented representation to reason about spatial direction information. His approach does only deal with point objects whereas we need to model two-dimensional objects like cars seen from the birds-eye perspective.
Mukerjee and Joe (1990) 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 direction. They then divide the plane around this box by extending the boxes' lines in eight two-dimensional regions that are named 1 to 8.
The direction relation matrix used by Goyal and Egenhofer (2001) 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. (2000). 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.

5. Describing an object constellation

When a constellation of several objects has to be described we can do this by stating each object's relation to the other objects. Often not all relations need to be given to get a full description of the scene. In figure 3a) you see three objects. Object 2 is right front of object 1 and object 3 is right front of object 2 and of course right front of object 1. The last relation however is implicit, given the other two relations. Object 3 cannot have another relation then right front of object 1 therefore the relation between object 1 and object 3 does not have to be mentioned explicitly. As you see here the order in which the relations are mentioned is important. If we say that object 2 is right front of object 1 and so is object 3, we still have to mention that object 3 is right front of object 2 and end up with giving three relations whereas the correct two would have sufficed. On the contrary in a situation like in figure 3b) you need to state all three relations: object 2 is straight front of object 1 and object 3 is right front/right neutral of object 1 as well as right neutral/right back of object 2.

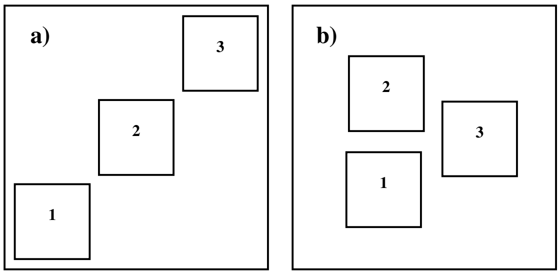

Figure 3 a) Object 3 is right front of object 1 is implicit in the information that object 2 is right front of object 1 and object 3 is right front of object 2 and therefore has not to be stated explicitly. b) Object 3 has a relation to the objects 1 and 2 that both have to be described to get a clear picture of the situation.

Altogether in a scene with n objects at most $\binom{n}{2}$ relations have to be given, when the directions of objects is the same for all objects or is not taken into account.
It is always possible to reconstruct a constellation of objects if the describer mentions each object first in relation to a reference object which must be one of the objects already in the reconstructed picture. The qualitative region of this reference object gives the first limitation of the space for the new object. In addition to that, some other objects, also already in the scene, will have an impact on where, within that qualitative region of the reference object, the new object has to be placed. We will call these objects the *influencing objects* to distinguish them from the first given reference object. Influencing objects are those

that can be find in qualitative regions of the reference objects that are horizontally or vertically *in-line* with the qualitative region of the reference object where the new object has to be placed.
That means for example, for an object that needs to be placed in the right front region of the reference object, that all relations between this object and the objects that are already placed in the areas right front, straight front, left front, right neutral and right back have to be given.
Regard the example in figure 4. The aim is to reconstruct the situation in figure 4a). The listener knows about object 1 already (figure 4b). He then learns that object 2 is straight front of object 1 and places it there (figure 4b again). The next information he gets is that object 3 is right front of object 1 which means that object 3 could be anywhere in the dotted area in figure 4c). To get the qualitative relations between all the objects right he needs to know which relation object 3 has to object 2. That means that the position of object 3 will be influenced by object 2. The listener can request the relation information and learns that object 3 is right neutral of object 2, which clarifies where object 3 has to be placed in order to get all qualitative relations between all objects right. (figure 4d)).

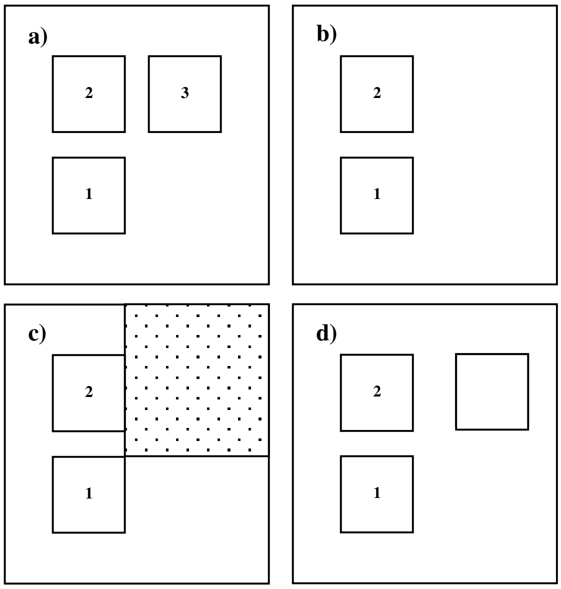

Figure 4 a) the original picture that has to be reconstructed form the listening agent. b) the listening agent knows already where object 1 is situated and can place object 2 after it is told that object 2 is straight front of object one. c) Object 3 shall be right front of object 1 that could be anywhere within the dotted area. The relation to object 2 is needed as well. d) The last relation (object 3 right neutral of object 2) is given and the object can be placed in the picture.

6. A mechanism to reconstruct an object constellation

6.1 The process

It is always possible to merge a new object into an existing picture. You first go to the qualitative region of the given reference object where the new object needs to be placed. Then you get the relations of all objects that are in a horizontal or vertical in-line region. According to those relations the space for the new object is narrowed down further and further. The order in which the relations of the influencing objects are given can further reduce the amount of relations needed. By cutting down the space for the new object with each learned relation, the number of influencing objects will often decrease. We will come back to this in section 6.3 additional remarks.
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). When moving objects, care has to be taken not to change existing relations.
We can solve this problem when we divide all the objects in the picture in two groups, one group containing all the objects that will be on one side of the new object and the other group containing all the objects that will be on the other side of it. Then all the objects that are on one side are moved outwards and new space appears for the new object. Moving the objects like this will not change any relation between them.

6.2 A simple example

The idea gets clearer if we consider the example in figure 5. Here all objects have the same direction, which means in this case that they are all facing north. That is of course not necessary but keeps the example simple at this moment. The object constellation in figure 5a) has to be reconstructed. The first object to be given is object 1. The listener does not know where the object is in exact coordinates and places it

somewhere. The first relation he learns is 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 5b).
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 reachable region. (Object 2 is right neutral of object 1.) Therefore the listener must know the relation between object 3 and object 2. It happens that object 3 is left front of object 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 5c) 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 5d). Free space appeared where object 3 can be inserted, which is done in figure 5e).
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. Free space appears where object 4 can be placed but all other relations of the other objects in the picture stay the same (figure 5g).
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 5h). In addition we need to draw a horizontal line beneath object 3 and move all the objects under this line downwards (figure 5i)). Figure 5j) shows the resulting space after we have moved the objects, finally in figure 5k) object 5 is placed and we have reconstructed the scene in figure 5a).

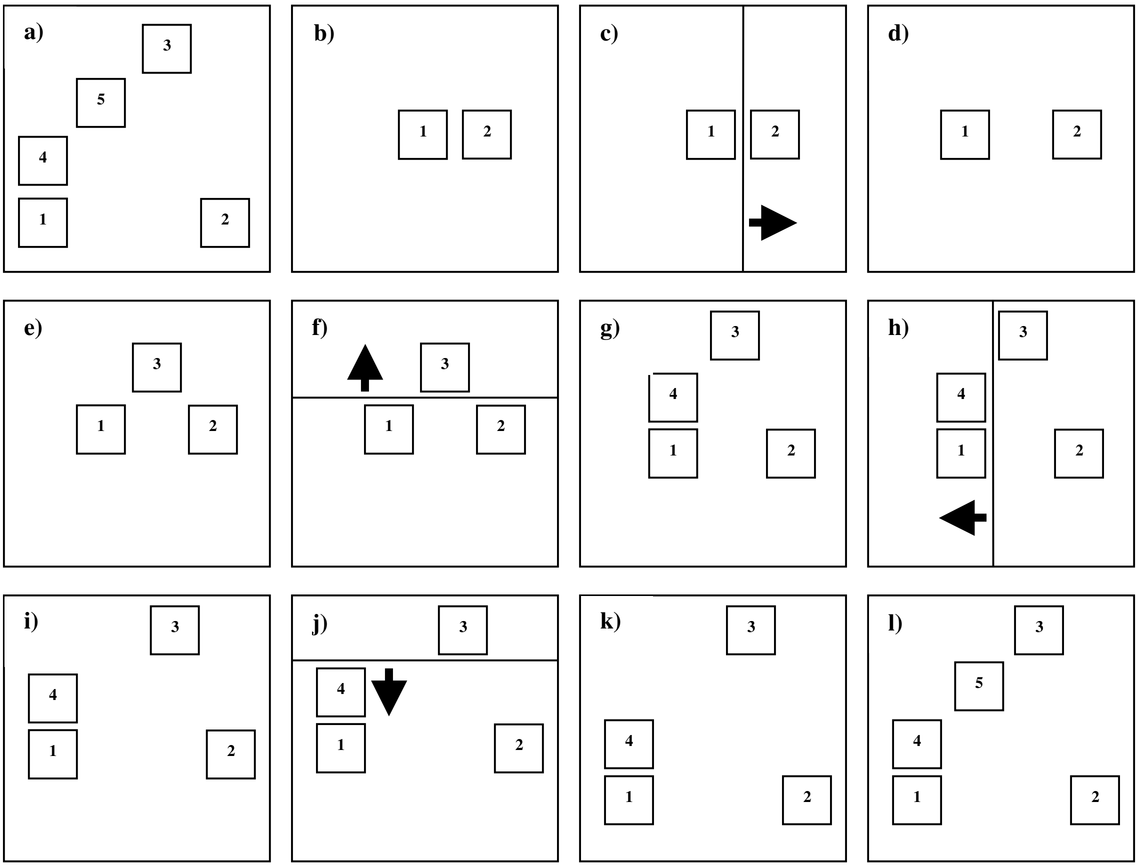

Figure 5 a) The original constellation that has to be reconstructed b) Object 1 is placed somewhere (it doesn't mater where because we are only interested in the qualitative position relations between the objects). Object 2 is placed right neutral of object 1. c) Object 3 has to be inserted between the objects 1 and 2 therefore object 2 is move outwards to the right. d) Free space appeared after moving object 2 outwards. e) Object 3 could 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 five needs to be right of the objects

1 and 4 therefore these objects are move 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 object 4 and object 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 therefore reconstructed. All relative position relations of the objects are the same as in the original in 5 a).

6.3 Additional remarks

Figure 5 illustrates just one example of how an object constellation can be reconstructed. There are several details that do not appear in the example but have to be mentioned in order to provide a clear understanding of how the whole process works.

- It is not necessary that the relative positions of the objects all are given to the same reference object first, like it was done in the example in figure 5. Any object that already is in the reconstructed picture can be given as reference object. Then you need to construct the qualitative regions around that object and after finding the right qualitative area to place the new object into, you need to check which other objects are influencing the new one. That are all objects in a region horizontally or vertically in-line with the region where the new object should be placed in.

- Furthermore in the example above all objects have the same direction, which is not necessary either. If we know about their direction we can use it. A direction can be described by stating both the relation R(a, b) and R(b, a). For each object we only need to give both relations once to obtain the direction. If we know object 1 is right front of object 2 and object 2 is right front of object 1, the direction of object 2 is clear. That means that when the scene contains n objects n additional relations have to be stated to obtain the objects' directions.

- We earlier said that all objects that are in horizontal or vertical in-line regions have an influence on the position of the new object. But the order in which these objects are taken into account can reduce the number of objects that have to be considered. Look at the example in figure 6: The objects 1, 2 and 3 are already known and placed in the picture. Now object 4 needs to be place in the right front region of object 1. That means so far that it can be anywhere in the dotted area in figure 6b). Object 3 and 4 are both in a horizontal in-line region (left front of object 1). If we start with the relation of object 4 to object 3 and we learn that object 4 is right front of object 3, the possible space for object 4 reduces to the intersection of the regions right front of object 1 and right front of object 3, shown in figure 6c). Now there is no further object that is horizontally or vertically in-line with the region where object 4 will have to be placed. Therefore the relation of object 4 to object 2 does not have to be given. In the future we will investigate further in which order the influencing objects should be taken into account in order to reduce as much effort as possible.

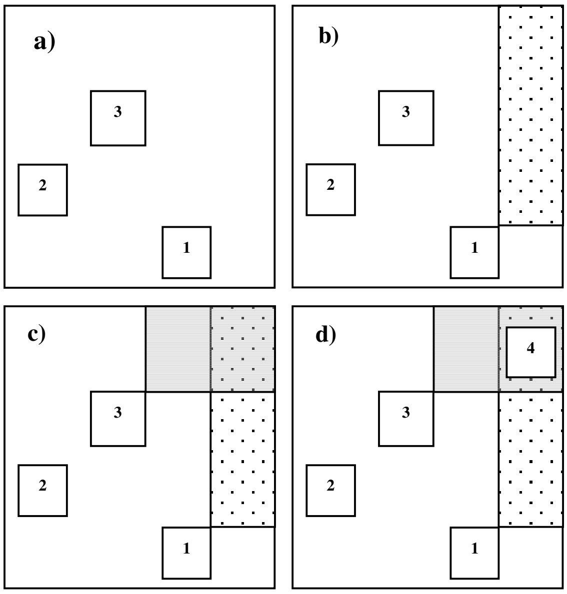

Figure 6 a) Three objects are already known and placed in the picture. b) Object 4 shall be right front of object 1, which is somewhere within the dotted area. c) The objects 2 and 3 are in a horizontal in-line region and therefore its relation to object 4 needs to be given. Object 4 is right front of object 3 which narrows the space for object 4 to the intersection of right front of object 1 and right front of object 3. Object 2 is not anymore within a region that is in-line with the space for object 4 and therefore its relation to object 4 is not needed. d) Object 4 is placed in the picture.

- It does not matter in which direction the objects are moved when some space has to be made. The relative positions will always be the same. That means each time we moved the object on the right side of a line further to the right we could as well have moved the objects on the left side of the line further to the left. The same, of course, is valid for moving upwards and downwards. Sometimes it can be necessary to move some of the objects twice. This was the case in figure 5h) and 5j) where we needed some space above and at the same time right of object 4.

- If the new objects that are merged into the picture are put as close as possible to the already existing objects the scene will develop nicely. At any time during the process the relations of the objects in the reconstructed picture will be the same as in the original scene.

7 Outlook

One application field of our research is vehicle traffic observation. This project is an offspring of the WITAS project (Doherty et al. 2000) which now deals with modeling a fleet of autonomous agents (helicopters, cars and human operators) that need to communicate about the traffic they observe. At first this should take place in a known environment where positions of houses and streets are common knowledge of all the agents. Later in the project we will explore even unknown territory where no landmarks are known so far.
The agents divide the area in quadrants and each of them observes one quadrant and tells the others what is going on there. In terms of the mental images we can interpret this in the way that each agent has a map of the whole territory. It fills in the traffic information for the quadrant it observes from the information of its own sensor data and for the other quadrants it learns what is going on there from the other agents' natural language descriptions.
The different types of agents do not only use different sensor techniques they also have different perspectives of the scene. The perspective of a helicopter is clearly different of the perspective of a car on the ground. An agent might even be able to change its own perspective. A helicopter can observe straight from above, watch the scene from an angle or even observe a part of the area while standing on the ground. Human agents can sit in a car or patrol on foot, and cars sometimes can be on bridges or mountains and oversee an area from above as well. Other agents might be at the mission control headquarters, not observing anything themselves and completely rely on the reports from the outside agents. All the different possible perspectives make it so important that the communication abstracts from them. If you think of an overtaking maneuver it does not have any impact on your interpretation that it is an overtaking maneuver, from what perspective you are observing it.
Our next step will be to implement the technique for reconstructing an object constellation with all qualitative relations preserved within the application of our project. We will investigate what order of influencing objects will turn out to be smart in order to reduce effort. The autonomous agents are going to use the technique for embedding the observations that they are told from the other agents and can not see them selves as well as they will be using it to produce well organized natural language descriptions of what they discover from their own sensors.

References

Doherty, P., Granlund, G., Kuchcinski, K. Sandewall, E., Nordberg, K., Skarman, E. and Wiklund, J. (2000). The WITAS Unmanned Aerial Vehicle Project. In: Proceedings of the 14th European Conference on Artificial Intelligence. IOS Press Amsterdam. 747-755.

Fernyhough, J., Cohn, A. G., and Hogg, D. (2000). Constructing qualitative event models automatically from video input. In: Image and Vision Computing, 18. 81-103.

Forbus, K. D. (1996). Qualitative Reasoning. In: Allen B. Tucker (ed.). The Computer Science and Engineering Handbook. CRC Press. 715-733.

Freksa, C. (1992). Using orientation information for qualitative spatial reasoning. In: A. U. Frank, I. Campari, U. Formentini (eds.). Theories and methods of spatio-temporal reasoning in geographic space. Springer, Berlin. 162-178.

Freksa, C. and Röhrig (1993). Dimensions of qualitative spatial reasoning. In: N. P. Carrete, M. G. Singh (Eds.), Qualitative reasoning and decision technologies, Proc. QUARDET93. 483–492.

Goyal, R. K., Egenhofer M. J. (2001). Similarity of Cardinal Directions. In: C. S. Jensen et al. (eds.). Advances in Spatial and Temporal Databases. SSTD 2001. LNCS 2121. Berlin Heidelberg: Springer-Verlag. 36-55.

Levinson, S. C. (1996). Frames of reference and Molyneaux's question: Cross-linguistic evidence. In: P. Bloom, M. Peterson, L. Nadel & M. Garrett (eds.), Language and Space. 109-169. MIT Press.

Mukerjee, A. and Joe, G. (1990). A qualitative model for space. In: Proceedings of the AAAI-90. The MIT Press Boston. 721-727.

Retz-Schmidt, G. (1988). Various Views on Spatial Prepositions. AI Magazine 9 (2). 95-105.

Sandewall, E., Doherty, P., Lemon, O., and Peters, S. (2003). Words at the Right Time: Real-Time Dialogues with the WITAS Unmanned Aerial Vehicle. In: Andreas Guenter et al (eds.). KI 2003: Proceedings of the 26th Annual German Conference on AI. Springer.

Steinhauer, H. Joe (2004). The Qualitative Description of Traffic Maneuvers. In: Workshop on Spatial and Temporal Reasoning. ECAI 2004.Valencia, Spain. 141-148.

Steinhauer, H. Joe (2004). A Qualitative Description of Traffic Maneuvers. In: Spatial Cognition2004: Poster Presentations. Frauenchiemsee, Germay. 39-42.

Steinhauer, H. Joe (to appear 2005). A Qualitative Model for Natural Language Communication about Traffic. In: Reasoning with Mental and External Diagrams: Computational Modeling and Spatial Assistance. AAAI 2005 Spring Symposium Series: Stanford University, California.

WITAS homepage http://www.ida.liu.se/ext/witas/