NATURAL LANGUAGE DIALOGUE ABOUT MOVING OBJECTS IN AN AUTOMATICALLY ANALYZED TRAFFIC SCENE

Heinz Marburger, Bernd Neumann
Hans-Joachim Novak

Fachbereich Informatik
Universitaet Hamburg
Schlueterstrasse 70
D-2000 Hamburg 13

ABSTRACT

This contribution is concerned with natural language dialogue about scenes with moving objects. Two systems are connected, a natural language dialogue system originally conceived for static scenes and an emerging scene analysis system for real-world TV-frame sequences. The latter produces time dependent object descriptions which serve as a referential database for inquiries. The time intervals relevant for answering the questions are determined from domain specific parameters, the context of the dialogue, the tense of the verbs and time adverbials. For checking the correspondence between a verbally specified motion and a trajectory, predicates are evaluated which can be deduced from the verb's case-frame.

I INTRODUCTION

The system described in this paper is designed to answer yes/no questions about moving objects in a recorded real-world scene. Scene analysis is performed independently up to a level of representation where each frame is symbolically described by object names linked to object types and associated with visible properties like position, shape and color. Identical objects in successive frames are recognized and given the same name. We call this level of representation, being scene dependent, referential knowledge.

Questions are asked in natural German language assuming the following pragmatic dialogue situation: we telephone with another person which is standing at a window, and ask questions about the traffic seen from this person's point of view. The natural language system tries to answer a question using the referential knowledge provided by the scene analysis system.

The architecture of the system has been strongly influenced by two independent investigations at the University of Hamburg. The scene analysis subsystem is being developed as part of a research effort towards understanding real world scenes with motion [1]. Current work concentrates on separating moving objects from static background and determining 3D-shape and trajectory of these objects. The dialogue subsystem is adapted from HAM-RPM [2,3] which works with a static world of discourse and has a large amount of linguistic capabilities at its disposal, e.g. pronoun resolution, handling of elliptical expressions, spatial relations, quantifiers and restrictive clauses as well as the capability to initiate clarification dialogues.

Both systems share a conceptual knowledge base which contains general knowledge common for a language understanding system, as well as information relevant for visually recognizing the objects in a scene, e.g. object shape descriptions. The referential database described earlier serves as the main communication channel between the two subsystems.

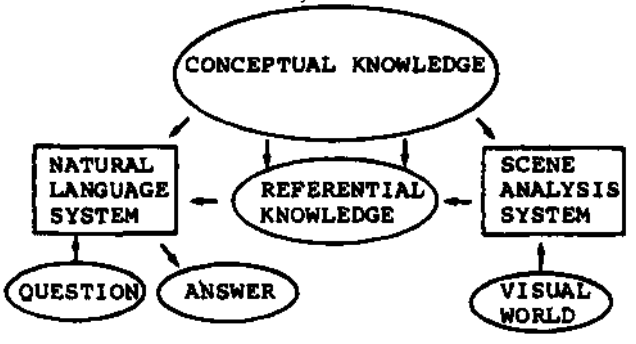

Figure 1: System overview

Our aim is to answer questions like (1) and (2).

(1) Hielt der Bus an?
(Did the bus stop?)

(2) Fuhr danach das gelbe Auto an?
(Did the yellow car start afterwards?)

We emphasize a top-down approach where verbalisations are processed in order to decide whether or not they properly describe a given image sequence. Most related research efforts differ from our approach in that they derive verbalisations bottom-up. For example the work of BADLER [4] which is further developed by TSOTSOS [5] offers motion conceptualisations which can be generated from image sequences and permit a crude verbalisation by simple translation of concepts into words. Similarily the systems of OKADA [6] or TSUJI ET.AL. [7] do not attempt to relate truly natural language expressions to the analyzed movements.

In this contribution we discuss three types of constraints which can be extracted from yes/no questions. They form the basis for a process which tests whether or not the symbolic scene representation corresponds to the inquiry. Verb tense and time adverbials give rise to a temporal constraint which is treated in the following section. The verb and its deep-case structure usually imply certain trajectory shapes. This includes the verb's manner of action which corresponds to trajectories satisfying particularly simple shape predicates. Finally, an example is given where location constraints can be derived from the deep-case structure.

II TENSE AND TIME

Dialogues in our system refer to a sequence of images covering a certain time span [Tbeg,Tend]. Within these limits a time of speech Ts can be arbitrarily chosen which synchronizes the dialogue with the scene. Ts is fixed throughout the dialogue, although some notion of progressing dialogue time is maintained for the purpose of modelling dialogue context. The system cannot access referential knowledge beyond Ts which marks the presence.

We call the subsequence of images which has to be regarded to answer a question, the (time) interval of consideration. It is mainly determined by the tense of the verb and possibly time adverbials, but also by the domain (i.e. its typical motions) and the course of the dialogue (i.e. the current focus of attention). We only consider the effect of tense and adverbials, and begin with tense. For present tense questions the boundaries of the interval of consideration are given by the parameters Tprl and Tpr2 which denote a short time span immediately preceding Ts. For present perfect and simple past the corresponding parameters are Tpal and Tpa2 which cover the scene from the beginning up to Tprl. Colloquial German does not necessitate a distinction between these tenses.

The effect of several time adverbials can be described as restricting the interval of consideration with respect to a time of reference. Let the event 'bus stops in question (1) terminate at Tbus2, then the time of consideration for question (2) extends from Tbus2 to Tpa2. Adverbials which refer to a time of reference which is defined by some other event correspond to the category ADVe introduced by BAEUERLE [8]. We have selected the following subset: 'vorher', 'davor', 'dann', 'nachher', 'spaeter', 'danach' ('previously', 'before this', 'then', 'afterwards', 'later', 'after that').

Other time adverbials specify the interval of consideration referring always to the time of speech as a special time of reference. They are summarized in the category ADVs. We only work with those adverbials of the category ADVs which put the interval of consideration close to the time of speech, namely 'jetzt', 'nun', 'gerade', 'gegenwaertig', 'im Moment', 'eben*, 'soeben' ('now', 'at present', 'just', 'just at the moment', 'just now'). This excludes adverbials like 'kuerzlich' ('recently*).

III LOCOMOTION VERBS

We investigate only questions involving verbs which denote a location change of the actor in the sentence. For a positive answer the trajectory of the actor, as recorded by the scene analysis system, must satisfy certain requirements or predicates which depend on the verb itself and the deep-case structure associated with it. For locomotion verbs we use the following case slots: AGENT, LOCATION, SOURCE, GOAL, PATH, OBJECTIVE.

Consider again question (1). Here the two primitive predicates 'moving' and 'stationary* have to be applied to the ACTOR trajectory over the interval of consideration. They can be easily computed from the object positions for each instance of the sequence. Four basic situations can be distinguished: intervals where the object is stationary, moving, beginning to move, and beginning to be stationary. The first case is not interesting since we are analyzing locomotion verbs. The other three possibilities represent verb inherent features, namely the manners of action: durative, inchoative and resultant. Some examples of verbs which can be analyzed using the above predicates are 'fahren' ('drive'), 'gehen' ('walk'), 'anfahren' ('start'), 'losgehen' ('start walking'), 'stoppen' ('stop').

Verbs like 'abbiegen' ('turn off) or 'wenden' ('turn') imply trajectories with more detailed properties. Predicates involving the change of direction of a trajectory will be required, none of which have as yet been accurately designed.

Finally, we consider the situation where slots are filled besides the AGENT slot. This may be due to prepositional clauses or transitive verbs.

(3) Bog das Auto von der Haltstrasse in die Fahrstrasse ab?
(Did the car turn off Haltstrasse into Fahrstrasse?)

In this example the slots SOURCE and GOAL are filled by 'von Haltstrasse' and 'in Fahrstrasse', respectively. The system not only checks whether there is a trajectory with the specified shape within the interval of consideration but also requires this trajectory to satisfy the spatial constraints imposed by the contents of the slots.

IV CONCLUSION

A natural language dialogue system and a scene analysis system for image sequences are being connected to explore the possibility of natural language communication with image understanding systems in general, and verbal description of motion in particular. A symbolic scene description in terms of time dependent object locations (and some additional properties) has been proposed as a level of representation suitable to serve as a referential database for inquiries. Answering yes/no questions about motion is viewed as a top-down process aiming at verifying certain trajectory properties. Three types of constraints on a trajectory can be distinguished. First, the interval of consideration as a temporal constraint, second, trajectory shape in space and time, third, spatial constraints on the location of a trajectory. It has been shown for some examples, how these constraints can be extracted from a question. The reported work is currently being implemented.

ACKNOWLEDGEMENTS

We wish to thank Walther v. Hahn, the head of the HAM-RPM project, for his helpful counsel in linguistic matters. We also gratefully acknowledge discussions with Hans-Hellmut Nagel who initiated this project.

REFERENCES

[1] Nagel, H.-H., Analyzing Sequences of TV-Frames, IJCAI-77, p.626, 1977.

[2] v.Hahn, W., Hoeppner, W., Jameson, A., Wahlster, W., The Anatomy of the Natural Language System HAM-RPM, in: Natural Language Based Computer Systems (Bole, ed.), Hanser Verlag, Muenchen, 1980.

[3] Jameson, A., Hoeppner, W., Wahlster, W., The Natural Language System HAM-RPM as a Hotel Manager: Some Representational Prerequisites, HAM-RPM Bericht 17, Germanisches Seminar, Universitaet Hamburg, 1980.

[4] Badler, N.I., Temporal Scene Analysis: Conceptual Descriptions of Object Movements, University of Toronto, TR-80, 1975.

[5] Tsotsos, J.K., A Prototype Motion Understanding System, University of Toronto, TR-93, 1976.

[6] Okada, N., SUPP: Understanding Moving Picture Patterns Based on Linguistic Knowledge, IJCAI-79, p.690-692, 1979.

[7] Tsuji, S., Morizono,A., Kuroda, S., Understanding a Simple Cartoon Film by a Computer Vision System, IJCAI-77, p.609-610, 1977.

[8] Baeuerle, R., Tempus, Adverb, temporale Frage, in: Wortstellung und Bedeutung, Akten des 11. Ling. Koll. Pavia, Bd. 1, Tuebingen, 1977.