Scene Visualization and Animation from Texts in a Virtual Environment

Fabrice Tabordet (*), Fabrice Pied (*) et Pierre Nugues (+)
(*) Université de Caen
(+) ISMRA
6, boulevard du Maréchal Juin
F-14050 Caen, France
pnugues@greyc.ismra.fr

Abstract

We describe here a prototype to recreate car accidents in a virtual world from written car accident reports. Reports have been supplied by the French MAIF insurance company and describe most often collisions between two vehicles.

We outline techniques that we used and results that we obtained. Although our project has not reached its end, we show that descriptions in French enable to animate scenes and to coordinate entities in a virtual world. We think that visualization and animation of the scene improve the intelligibility of the report.

Cognitive Value of Virtual Environments

Virtual reality techniques enable to recreate objects, scenes, to simulate actions, to reproduce scientific experiments with the aim to get as close as possible of reality. Virtual reality is now used in an increasing number of applications. They range from computer assisted object design, simulation of assembly plants to assess their efficiency, work group environments for virtual meetings and teleconference (Andersson et al. 1994).

Central features of virtual reality rest on image and interaction. Interaction eliminates the impression of a possible passiveness that may tire a user. It gives him/her a greater freedom in the exploration and navigation in its environment. Images, most often in a 3-D space, help a user understand a situation (Kosslyn 1983, Denis 1991). Visualization gives virtual environments the capacity to represent and communicate knowledge. In collaborative virtual environments, for instance, designers use object metaphors to help a user immediately access the sense of a scene, an artefact. Visualization facilitates the comprehension of complex scenes and data. Michel Denis (1996) notices

> *l'image et le langage constituent deux modes de représentations […] fortement différenciées, mais dont la coopération est requise dans de nombreuses formes de fonctionnement cognitif.*
>
> Image and language constitutes two representation modes […] strongly different, but requiring a cooperation from each other in many instances of cognitive activity.

Language and Interaction in a Virtual Environment

In virtual environments, interaction is nearly always carried out using pointing devices such as mice, space balls, joysticks, etc. These devices enable navigation and motions in vertical and horizontal planes and rotations. Such devices also enable to designate an object and to fly to it. Finally, pointers allow to control objects of a virtual world: to manipulate them, to activate them, etc.

Interaction in a virtual world is tricky. It's the counterpart of the better intelligibility. Classical pointing systems allow a fast and accurate designation provided that objects are in the vision field of the user. Navigation is relatively easy when the motion is rectilinear. It's more difficult when the motion is complex or when more artifacts have to be synchronized.

Natural language or spoken interaction is seldom used in virtual environments. In addition to being more "natural", language enables the user to carry out actions that would have been very difficult with the interaction paradigms common in virtual reality environments:

- direct interaction – objects should be visible
- gesture interaction – whatever the sophistication of the devices you are using, you have to do it by hand

One possible explanation is that reliable speech recognition devices are still scarce and that processing languages is difficult. However natural language control shows its superiority in some cases:

- *when navigating into virtual worlds*. Although, gesture interaction enables a user to carry out linear motions easily or to move quickly to visible objects, it gets more difficult when motions are complex and in reference to objects. In experiments we have conducted previously, we realized that novice users couldn't even move around a simple table without extensive training (Godéreaux et al. 1996);
- *in the designation of objects*. When not visible, objects are sometimes difficult to find in a virtual world. On the contrary, natural language search or simply using keywords is often preferred. Examples of this preference can be given by the popularity of the search robots for the Web. The situation is even more difficult when designating complex objects such as plurals;
- *when controlling an artifact*. Controlling an artifact in a virtual environment leads often to direct manipulations again. Some of them might still require complex scripts of commands and virtual reality techniques do not show obvious improvements over classical (and tedious) Unix shells. On the contrary, natural language commands even when constrained by the present state of technology can be easier and more efficient;
- *when delegating a task to one or more artifacts*. Delegating an action to an artifact can be considered as a complex control taking into account parameters such as time or space or conditional facts. Delegating also requires sometimes that events be synchronized from different types of source. Although processing is difficult, delegation is more easily encapsulated using language.

In a previous prototype, one of the author has addressed navigation and designation problems (Bersot et al. 1998). This prototype – Ulysse – consisted in a conversational agent that was embodied in the representation of a user in the virtual world. The user addresses the agent using spoken natural language asking it to carry out motion commands. Ulysse could then move the user in the virtual world on his/her behalf.

In this paper we describe a somewhat different application of natural language possibilities to outline cooperation principles between moving entities in a virtual world. We took the application domain of car accident reports. Reports were provided by the MAIF insurance company. The corpus consists of 89 reports written in French. They include descriptions such as this one:

Je roulais sur la partie droite de la chaussée quand un véhicule arrivant en face dans le virage a été complètement déporté. Serrant à droite au maximum, je n'ai pu éviter la voiture qui arrivait à grande vitesse. (Report A8)

I was driving on the right-hand side of the road when a vehicle coming in front of me in the bend skidded completely. Moving to the right of the lane as far as I could; I couldn't avoid the car that was coming very fast.

The interpretation of such reports is difficult because they contain understatements, negations, or cosmetic descriptions that cannot be interpreted without guessing the state of mind of the driver. In order to turn the processing tractable on a first attempt, we simplified some reports. Considering the previous text, the simplification is:

Je roulais sur la chaussée, un véhicule arrivait en face dans le virage, je l'ai percuté dans le virage.

I was driving on the lane, a vehicle was coming in front of me in the bend, I bumped it in the bend.

Our goal is to generate and animate the corresponding scenes in a virtual world, using texts and knowledge of driving conditions in France. A text may be considered as a guide to re-create the scene, and not as a complete scene description: the author of the text will not give all that is needed to re-create the scene in question, and this activity needs a great deal of common knowledge about the domain under consideration. Considering for example the first proposition of the text A8, the writer assumes from the outset that we know he is driving in a correct and normal fashion (following the course of the road, at a correct speed).

In order to reproduce this reasoning, we used a processing architecture consisting of linguistic analyzer and a road domain model. This architecture seems sufficient to reproduce relatively simple events. We also implemented a planner to reproduce complex motions events occurring in the reports.

Semantic Analysis

The first stage of the processing consists in a morphological and syntactic analysis, splitting the text into segments, such as noun groups or verb groups, and establishing links between these segments (Poirier 1996). This step was carried out using a dependency analysis (Tesnière 1959, Mel'cuk 1988). Then the verb valence structure is computed to the corresponding meaning. This analysis uses a lexicon containing for each verb, the possible valence structures and the corresponding representation. Other units identify entities and link them together using space and time relations. Finally, linguistic processing splits sentences into a sequence of events, and the result is a set of entities and the relations that link them. We give a brief overview of the results we have obtained so far with the processing of the text A8:

I was driving on the lane

- Entities:
  `v0: vehicle; s0: person; r0: road; e1: trajectory; im1, ip1, ie: intervals`
- Time relations:
  `contains(ip1, im1) & before(im1, ie)`
- Relations:
  ```
  [ip1] driver(s0, v0)
  [ip1] within(v0, r0)
  ```

a vehicle was coming in front of me in the bend,

- Entities:
```
v1: vehicle; s1: person; sr0: bend; e2: trajectory; im2,
ip2, ie: intervals
```
- Time relations:
```
contains(ip2, im2) & before(im2, ie) & simultaneous(im1,
im2)
```
- Relations:
```
[ip2] driver(s1, v1)
[ip2] within(v1, rr0)
[ip2] getting_closer(v1, v0)
[ip2] facing(v1, v0)
```

I bumped it in the bend.

- Entities:
```
v1: vehicle; s0: person; e3: bump; im3, ip3, ie: intervals
```
- Time relations:
```
contains(ip3, im3) & before(im3, ie) & before(im2, im3)
```

For the first set of relations, corresponding to *I was driving on the lane*, we introduce in the representation a person `s0` and his vehicle `v0`, a road `r0`, an event `e1` corresponding to the movement of the driver, and also some time intervals (`im, ie, ip`) which allow us to situate the events in a chronological order and to establish if they have finished or not. These results are detailed in Pied (1996) and they rely on Gosselin's theory (1995) on tense and aspect of French verbs.

We have also some relations specifying that the driver of the car `v0` is `s0`, and that, during the process, the driver is on the road (`[ip0] within(v0, r0)` with `[ip0]` designating the length of the event).

These results are integrated, one after the other, in the scene construction unit and they are combined with some knowledge of the domain, in order to complete the description. Then, the reasoning processes begins and include a construction of a chronology of events and a construction of the vehicles' trajectories. The description is then translated into the VRML geometric formalism. It features a modeling of cars, roads and "road events" and takes as an input:

- each event with its temporal length
- the list of objects to draw
- their positions, speed and directions at the beginning and at the end of each process.

Scene Synthesis

The scene synthesiser uses the entities determined by the semantic module. The synthesizer builds the geographic landscape using standard road components, namely for the moment, straight roads, bends, and junctions. It also uses pre-defined vehicles such as the vehicle in Fig 1.

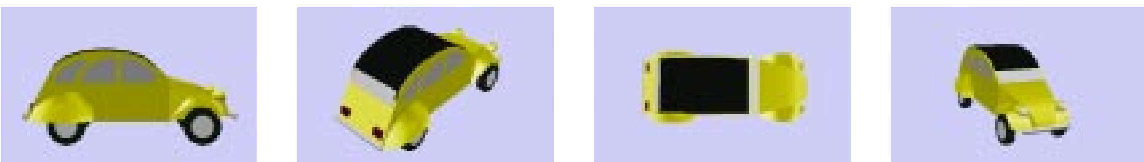

Figure 1 Our car model.

Road components are assembled respecting realism constraints, that is road world plausibility. The program sets the vehicles on the road and uses time and space relations to compute their positions split into the intervals. We have implemented the animation with VRML-2 interpolators. The scene corresponding to the previous report is shown in Figure 2.

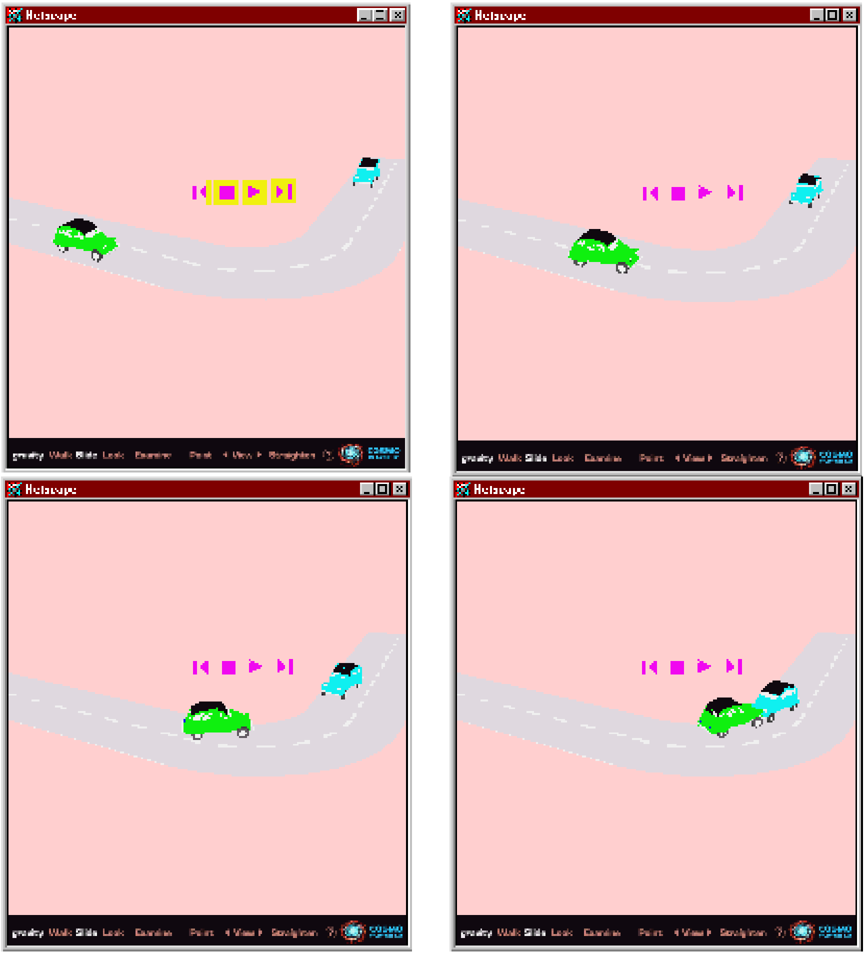

Figure 2 The synthesized scene.

Planning

One of the main problem in reproducing car accidents is to formalize information relative to driving events. Some complex driving events are difficult to model using the scene representation module we described above. For instance, it is impossible to encode the moving position of a vehicle relatively to the road it is on and at the same time, relative to the position of another moving vehicle. In order to address this, we selected a couple of actions such as overtaking, or behavior of vehicles at a junction and we implemented planning rules to reproduce them.

Making a plan consists in constructing a sequence of actions to solve a problem (Fikes and Nilsson 1971), these actions being sequential or parallel. In addition, planning should be reactive (Wilkins et al. 1995), that is actions cannot always be pre-computed and sometimes depend on the current state of the geometrical database. For instance, representing an overtaking scene requires splitting a complex action into a set of simpler ones. It also requires that the overtaking vehicle constantly monitors the position of the vehicle that is overtaken.

Our planner uses a hierarchical decomposition of a task into sub-tasks (Sacerdoti 1974) that we illustrate here on the overtaking action. The planner is embedded in the following vehicle which is called the "actor". This vehicle updates its position relative to the vehicle in front: the "object". The planner applies rules to compute the successive positions and orientations of the vehicles within the time frame in which the action occurs.

A planning rule is a predicate composed of:

- calling arguments;
- one or more actions to execute (for example, a position or an orientation calculation);
- the stop condition that defines the triggering of a subsequent rule;
- a call to the following rule, when the stop condition is satisfied.

The initial problem *overtake* is decomposed into sub-problems. It must fulfill an initial condition: "go faster than the other vehicle". Then it must complete a sequence of actions:

- get closer to the leading vehicle;
- move to the left lane;
- go in front of the overtaken vehicle;
- finally return to right line.

The rules are triggered by comparing positions of the different vehicles. For instance, the actor changes lane when reaching a given distance from the object. The overtaking action is represented by an action graph in Figure 3.

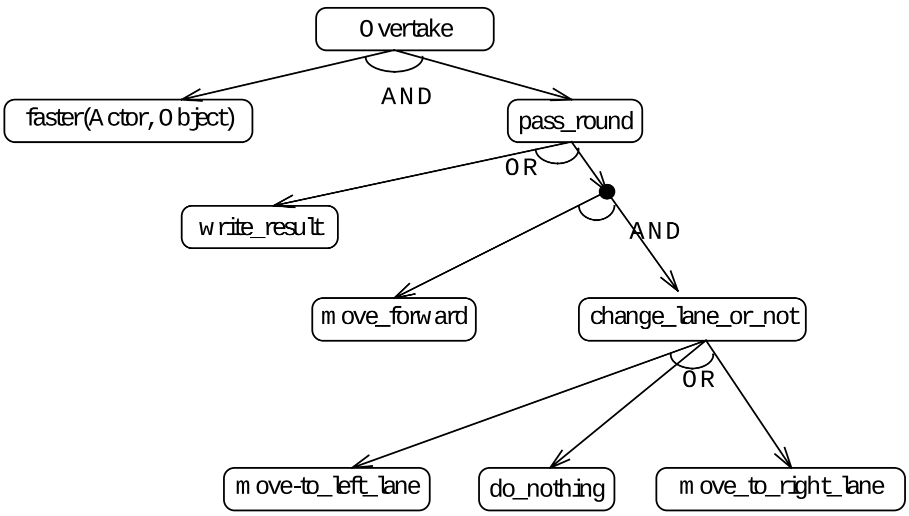

Figure 3 Graph of actions.

Motions of both vehicles are computed relative to the coordinates of the road reference system. The program computes in parallel one plan for each vehicle. Knowing the positions and the entities of the scene, the system compiles automatically the interpolators that define the overall motion. Figure 4 shows snapshots of the synthesized overtaking.

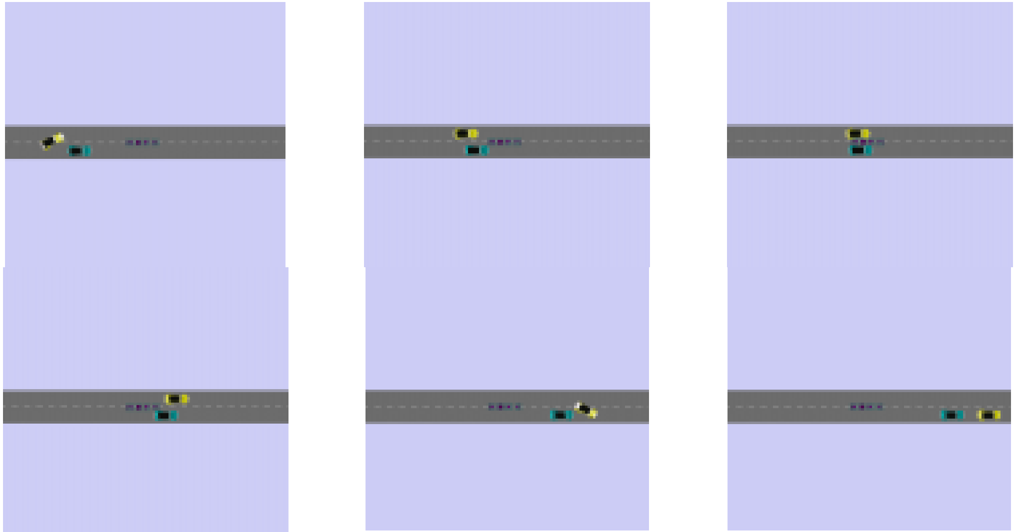

Figure 4 Overtaking.

Conclusion

We have described here techniques to synthesize and synchronize events from written texts. Such scenes could not have been animated by hand in VR environment using classical interaction techniques. Although texts need to be simplified and the number of actions is still limited, we have obtained some promising results.

In the future we plan to expand the semantic parser so that it can consider non simplified reports. We will also enrich our set of road configurations and scenarios to other road events. For the moment, they are limited to collisions on straight roads or corners, and to planning overtaking and junction behavior. We also plan to register the accident scenes into a 3D geographical database that is now available for most parts of France.

References

M. Andersson, C. Carlsson, O Hagsand, O. Ståhl, *DIVE The Distributed Interactive Virtual Environment. Technical Reference*, Kista: Swedish Institute for Computer Science, 1994.

Ball, G., Ling D., Kurlander D., Miller J., Pugh D., Skelly T., Stankosky A., Thiel D., Van Dantzich M., & Wax T.: Likelike Computer Characters: The Persona Project at Microsoft Research, in *Software Agents*, J. Bradshaw ed., MIT Press, 1997.

O. Bersot, P.O. El Guedj, C. Godéreaux, & P. Nugues, A Conversational Agent to Help Navigation and Collaboration in Virtual Words, *Virtual Reality*, 3:71-82, 1998.

Bolt, R. A.: Put That There: Voice and Gesture at the Graphic Interface, *Computer Graphics*, vol. 14(3), pp. 262-270, 1980.

M. Denis, Imagery and Thinking, In *Imagery and Cognition*, C. Cornoldi and MA McDaniel eds. Springer Verlag, 1991.

M. Denis, *Rapport scientifique du Groupe Cognition humaine*, LIMSI, Orsay, 1996.

Everett S., Wauchope K., Perez M.A.: A Natural Language Interface for Virtual Reality Systems, *Technical Report of the Navy Center for Artificial Intelligence*, US Naval Research Laboratory, Washington DC, 1995.

R. Fikes & N.J. Nilsson, A new approach to the application proving to problem solving, *Artificial Intelligence*, 2:189-208, 1971.

C. Godéreaux, K. Diebel, P.O. El Guedj, F. Revolta, & P. Nugues, An Interactive, Spoken Dialog Interface in Virtual Worlds, in: *Linguistic Concepts and Methods in CSCW*, J.H. Connolly & L. Pemberton, Chapter 13, pp. 177-200, Springer, London, 1996.

L. Gosselin, *Sémantique de la temporalité en français, un modèle calculatoire et cognitif.* Habilitation à diriger des recherches, Université de Caen , 1995.

Karlgren, J., Bretan, I., Frost, N. & Jonsson, L.: Interaction Models, Reference, and Interactivity in *Speech Interfaces to Virtual Environments, 2nd Eurographics Workshop*, Monte Carlo, Darmstadt, Fraunhofer IGD, 1995.

SM Kosslyn, *Ghosts in the Mind's Machine: Creating and Using Images in the Brain*, New York, WW Norton, 1983.

I.A. Mel'cuk, *Dependency Syntax: Theory and Practice*, State University of New York, Press 1988.

F. Pied, C. Poirier, P. Enjalbert, & B. Victorri, From language to model, *Proc. of ECAI-96*, 1996.

C. Poirier, *Conception et réalisation d'un analyseur sémantique de documents composites (textes et schémas)*, Thèse de doctorat, Université de Caen, 1996.

E. Sacerdoti, Planning in a hierarchy of abstraction spaces, *Artificial Intelligence*, 7:231-272, 1974.

L. Tesnière, *Éléments de syntaxe structurale*, Klincksieck, 1959.

D. Wilkins, K. Myers, J. Lowrance, & L. Wesley, Planning and reacting in uncertain and dynamic environments, *Journal of Experimental and Theoretical AI*, 7(1):197-227, 1995.