Interactive Environments for the Exploration of Large Data Spaces

Robert G. Belleman [1], Jaap A. Kaandorp and Peter M.A. Sloot

Parallel Scientific Computing and Simulation Group,
faculty of Mathematics, Computer Science, Physics, and Astronomy,
University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, the Netherlands.

Abstract

With the development of high-performance computing techniques and the increase in the availability of computer power, large and complex time dependent data sets are generated in simulations used in industrial and scientific applications. These data sets are not only large, but they also represent simulations of increasingly complex phenomena which often vary dynamically. In many cases, visual exploration of these complex data sets is one of the few options to analyze this data and to obtain further insight in the simulated phenomena.

As an initial test case in the development of an immersive virtual exploration environment we have carried out simulations of biological growth processes. An immersive virtual environment such as the CAVE aids in the interactive exploration of the large scale and time dependent data sets that result from these simulations.

Key words: interactive data exploration, online simulation.

---

1 Introduction

The increase in availability of computational resources, both storage capacity and processing speed, have allowed researchers in industrial and scientific areas to investigate increasingly large and complex time dependent problems. As a result, the data sets that are generated by these applications grow larger and more complex. Furthermore, many of the industrial and scientific applications are typically simulations of complex systems. In general these problems are intractable and NP complete so that the only available option to obtain insight in these problems is through explicit simulation and by doing a guided

---

[1] Contact: `robbel@wins.uva.nl`

Preprint submitted to Elsevier Preprint *24 February 1999*

search through the problem's phase space [5]. As these phase spaces grow larger, the time that is required to find a satisfactory solution often becomes unacceptable.

In general there are two types of experiments researchers perform with simulation systems which we will here refer to as (1) the *batch* simulation model and (2) the *living* simulation model. In the *batch* simulation model the parameters for the simulation are set and the application is run without any further intervention. The results are obtained after completion of the application which in some situations may take many hours to days. As the data sets containing the results grow larger and more complex, the demands that are posed on data analysis tools increasingly exceed their capabilities. In these cases researchers have to resolve to other means of analysis, in most cases some form of data presentation such as visualization, from which researchers can deduce information using their personal expertise and cognitive abilities.

In the *living* simulation model the application runs while the researcher interactively alters parameters that influence the course of the simulation. This mode of execution allows researchers to explore the phase space of a simulation *during* its execution, effectively taking part in the simulation, possibly directing the simulation towards an acceptable solution in less than the time required by the algorithms alone.

With both simulation models, the size and complexity of the problem spaces that need to be explored require a high performance presentation environment that immerses a researcher into the problem space. Such an exploration environment should provide a sufficient level of fidelity and interaction to encourage the researcher to explore the presented data, allowing meaningful tasks to be performed. Present immersive virtual environments are now just reaching the requirements to fulfill these high standards. Within the context of a recently initiated project at the Parallel Scientific Computing and Simulation group we will describe the issues involved in the design of such an interactive virtual exploration environment, and we will present our experiences with a rudimentary system for the interactive exploration of simulated biological growth systems that we have recently built and experimented with.

2 Interactive exploration environments

It is our objective to investigate the different issues that are involved in the development of an interactive exploration environment that allows a meaningful exploration of large data and phase spaces. We distinguish three different components in such a system: the available technology, the applications and the scientific issues.

2.1 Technological issues

The technological aspects are concerned with the computational hardware that is used for the exploration system. Although the capabilities of modern computer systems are nearing the requirements for performing both the exploration and presentation tasks on the same machine, some performance increase may be attained by running these tasks on dedicated machinery. For example, many simulation applications perform better on dedicated hardware such as vector processors, massively parallel platforms or other high performance computing machinery, while state-of-the-art graphical systems are now available that are well suited for the presentation tasks. Where the simulation and exploration tasks are distributed over different systems, some means of coupling is required between the two (see Figure 1).

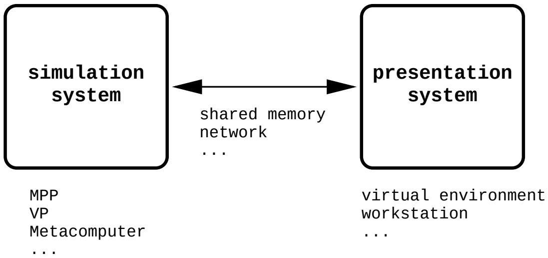

Fig. 1. Schematic configuration of an interactive exploration system.

An important goal towards a successful exploration environment is to involve the researcher into the presentation as much as possible, thereby increasing the researcher's level of awareness. To achieve this, an exploration system needs the following, often conflicting capabilities:

- Quality presentation – The most common method to provide insight in a simulated phenomenon is to represent the abstract data as visual geometric constructs that present quantitative and relational aspects to the observer in an intuitive manner. Over the last decade, many visualization techniques have been developed that allow data sets to be presented efficiently and in an intuitive manner. However, open issues remain on ways of incorporating data sets that are generated by the various multi-disciplinary research areas [1]. Furthermore, in the applications we want to use we know that mere visual exploration is not sufficient. For this reason we intend to integrate additional sensoral components for the exploration system, such as aural and haptic feedback, thereby increasing the user's field of perception.
- Rapid frame rate – While the capabilities of modern graphical workstations allow the construction of high quality and complex images with relative ease, the level of detail in the presentation should be minimized to avoid information clutter and achieve high frame rates (the aim should be to employ "minimal means for maximum effect"). For a usable exploration environment the visual frame rate should be at least 10 frames per second.

- Intuitive interaction – Some level of interaction with a presentation is mandatory. Unlike the standard interaction metaphors used in windowing systems on workstations, no standard user interaction metaphors yet exist for virtual environments. The capabilities of currently available virtual environments now allow researchers to interact with stereoscopic graphical presentations using three-dimensional input devices which provide a more-or-less intuitive interaction paradigm.
- Real-time feedback – Some delay will always occur between the moment a user interacts with a presentation and the moment that a response is available. This is caused by low tracking rates of input devices, communication delay between the exploration and the simulation system and temporary reduced availability of computational or network resources. To attain accurate control over a running application the amount of lag in an exploration system should not exceed 0.1 seconds.

2.2 Application issues

Each application has different execution characteristics with respect to the update frequency, the point in the application where interaction may occur and at what moment, and the amount of data that changes inbetween each update and thus needs to be communicated to the presentation system.

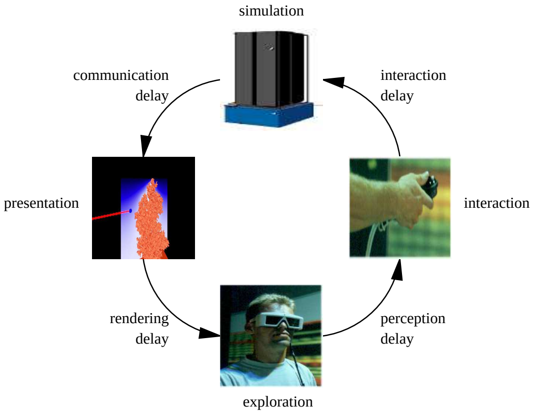

Fig. 2. Schematic representation of a typical interaction loop.

The relative importance of each of these capabilities depend primarily on the characteristics of the simulation application and the limits imposed by the infrastructure that is available. It is our intention to develop a conceptual model with which we can simulate the behaviour of an exploration system before it is built. To obtain an accurate representation of the behaviour of the system as a whole, this simulation should include the temporal characteristics of the simulation application, the presentation software and the interconnection between the two.

2.3 Scientific issues

The key issue towards a successful exploration environment is real-time interaction [2]. One way to achieve this is through a trade-off between speed and accuracy of the exploration by allowing “short-cuts” to be made in in a guided search by an interactive feedback mechanism between the application and the exploration system. The scientific question we want to address is concerned with the implications of making these “short cuts”.

The guided search algorithms that are used to search a problem’s phase space are often based on methods that stem from mathematical, biological or physical models. These macroscopic models represent abstractions of search methods that have been observed in nature. To be able to perform computer simulations using these models, the macroscopic models are converted into a computable algorithm. At this microscopic level, it is expected that the computable algorithm represents a sufficiently accurate abstraction from the macroscopic model such that it achieves the desired result. However, the process of making “short cuts” on a microscopic level may have severe implications on the macroscopic model that is being simulated. Indeed, the resulting simulation may well be totally different from that described at the macroscopic level.

3 A preliminary test case: simulated biological growth

An example of an intractable problem which we have used for our initial experiments towards the development of an exploration environment is the following.

3.1 Background

In the development of many biological systems, the distribution of chemical agents and nutrients plays a fundamental role. For filter-feeding marine sessile organisms, such as stony-corals, the growth process is affected by the distribution of suspended material in the external environment. From the biological literature it is well-known that water movement may have a strong impact on the shape of stony-corals. It is often possible to correlate growth forms of stony-corals with the amount of water movement. Compact growth forms are generally found under conditions with a large exposure to water movement, while the growth form changes gradually into a branching shape when the amount of water movement decreases.

In Figures 3 and 4 two growth forms of the stony-coral species *Pocillopora damicornis* are shown. The compact form in Figure 3 originates from an exposed site and the thin-branching form in Figure 4 was collected from a sheltered site.

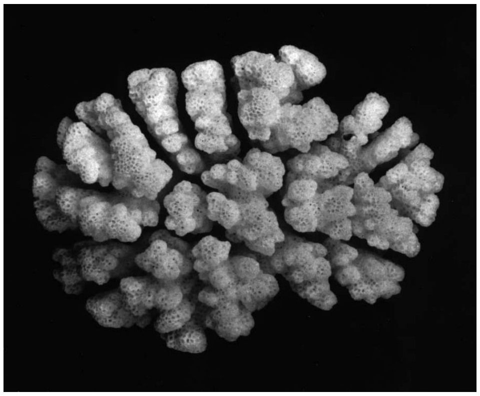

Fig. 3. Growth form of the stony-coral *Pocillopora damicornis* originating from an exposed (to water movement) site.

Fig. 4. Growth form of the stony-coral *Pocillopora damicornis* originating from a sheltered site.

3.2 The model

We have studied the effect of hydrodynamics on a very simple type of growth process, viz. growth by aggregation. The basic construction we have applied in the simulation is shown in Figure 5. In the model, aggregation proceeds by the accumulation of a "nutrient". The nutrient distribution is modeled using a Lattice Boltzmann model of transport. The aggregate absorbs the nutrient and the amount absorbed determines the local growth probability. In Figure 5 the flow is directed from left to right, the nutrient is released from the top plane and in each growth step a site is added to the aggregate.

3.3 Experiments

We have carried out simulations of growth processes (aggregation processes) in which an aggregate consumes nutrients from its environment and where nutrients are dispersed by a combined process of flow and diffusion [4]. The effect on the aggregate caused by different rates of fluid flow and nutrient

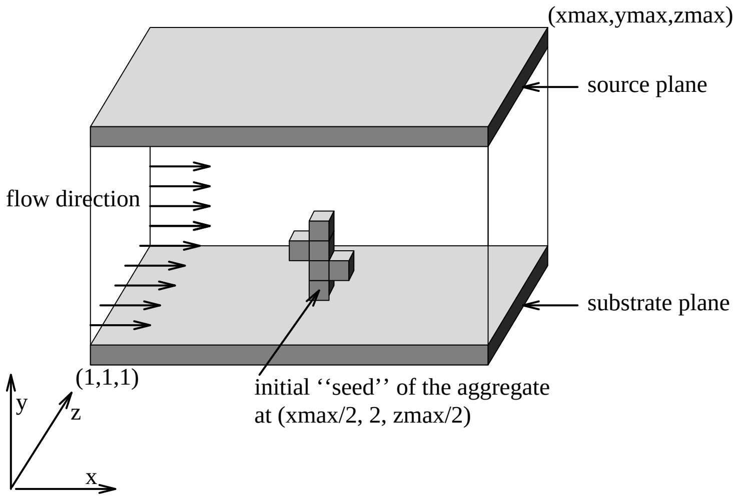

Fig. 5. Basic construction of the aggregate.

dispersion is investigated. The data resulting from these simulations includes the growth of the aggregate over time and the dispersion of nutrients around the aggregate.

In batch execution mode, this type of simulation runs for 72 hours on a 16 node 2 GFLOPs Parsytec system. The only method to obtain insight in the results of the simulation is through visualization of the generated data sets and visually comparing these with existing coral structures. A special purpose software package was implemented to obtain surface models of the generated structures at each growth step which are then visualized on a graphical workstation. The complexity of the aggregates are such that the generated surface models are too big to allow interactive exploration. In such cases it is often necessary to generate animations on video which take well over a week to produce. In addition, the end results are inherently non-interactive which impedes exhaustive exploration.

3.4 Data exploration in the CAVE

We have built an interactive exploration system that allows the data sets that are generated by the simulation to be explored inside a CAVE [3] located at the Academic Computer Services Amsterdam (SARA). We expected that a virtual visualization environment like the CAVE aids in the exploration of the large scale and time dependent data sets that result from these simulations.

In Figure 6 the result is shown of a simulated aggregation process. The aggregate emerges in an environment where nutrients are mainly dispersed by diffusion. In this case an irregular branching aggregate is formed. The nutrient distribution around the aggregate is visualized using a blue-white gradient, where blue indicates a high and white a low nutrient concentration. In these experiments a similar trend as observed in the stony-corals is demonstrated: as soon as the influence of hydrodynamics increases a more compact aggregate emerges.

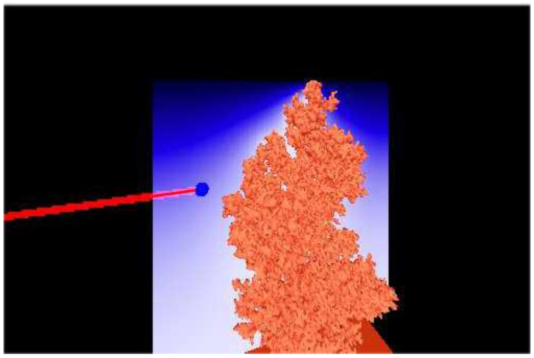

Fig. 6. CAVE application for the exploration of aggregation processes.

Using the exploration environment in the CAVE we have investigated simulated results of various experiments in which we have varied the influence of hydrodynamics on the growth process. In addition, we have been able to compare these results with CT scans of actual stony-corals. Using these CT scans, a more flexible comparison of the simulated structures to the structures found in nature is now possible (see Figure 7).

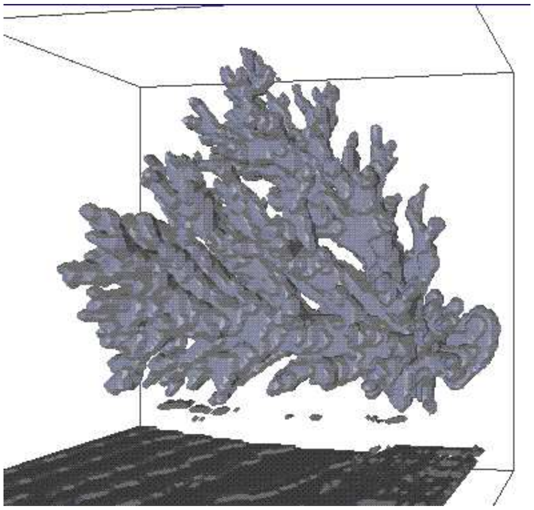

Fig. 7. Surface visualization of a CT scanned coral structure.

Our main finding is that the CAVE allowed us to study the effect of flow on the nutrient distribution far easier than was previously possible. The exploration environment allowed us to study the geometric structure of the aggregates simply by walking around the presented object. Using the CAVE's simple to operate "wand" (a 6 degrees of freedom flying mouse) we were able to easily explore the growth of the aggregate over time, and the dispersion of nutrients around the aggregate.

4 Future work

The growth simulations described in the previous section are currently performed off-line, using a parallel version of the software executing on a Parsytec CC-40 and CC-16 which are both located at SARA. In the near future we intend to develop an on-line version of this model which will be closely coupled to the exploration environment in the CAVE. Such an environment should allow the growth to be monitored and influenced during simulations.

Following that, we will investigate the incorporation of haptic feedback technology inside the SARA CAVE for determining the orientation of a CT scanned coral structure with respect to the flow direction in which a maximum of nutrient can be captured. In principal this orientation can be computed, however, given the fact that the number of possible orientations of such a structure is huge, it is our belief that through haptic feedback technology we will be able to exploit personal expertise of the investigator through additional sensory devices.

5 Acknowledgements

We gratefully acknowledge the work of Harro Mantel in the development of the CAVE exploration environment. We would also like to thank the people at SARA for the opportunity and their assistance in its development.

Thanks to dr R.W.M. van Soest of the Department of Coelenterates and Porifera, Institute for Systematics and Population Biology, University of Amsterdam, for allowing us to use collection material for generating the CT scans.

Thanks to Dick Bakker of the Academic Hospital Leiden, the Netherlands, for making the CT scans.

References

[1] Steve Bryson. The next ten years of visualization. *Computers in Physics*, 11(4):362–369, Jul/Aug 1997.

[2] Steve Bryson and Sandy Johan. Time management, simultaneity and time-critical computation in interactive unsteady visualization environments. In

Proceedings of Visualization '96, page 255. IEEE Computer Science Press, Los Alamitos, CA, 1996.

[3] C. Cruz-Neira, D.J. Sandin, and T.A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *SIGGRAPH '93 Computer Graphics Conference*, pages 135–142. ACM SIGGRAPH, August 1993.

[4] J.A. Kaandorp, C. Lowe, D. Frenkel, and P.M.A. Sloot. The effect of nutrient diffusion and flow on coral morphology. *Physical Review Letters*, 77(11):2328–2331, 1996.

[5] P.M.A. Sloot, J.A. Kaandorp, and A. Schoneveld. Dynamical Complex Systems (DCS): A new approach to parallel computing in computational physics. Technical report, University of Amsterdam, 1995. Technical Report FWI TR: CS-95-08.