Proceedings of the 2000 IEEE
International Conference on Robotics & Automation
San Francisco, CA • April 2000

Design Constraints for Haptic Surgery Simulation

Oliver R. Astley and Vincent Hayward
Department of Electrical Engineering and Center for Intelligent Machines
McGill University
3480 University Street, Montréal, Qc, H3A 2A7, Canada
oliver@cim.mcgill.ca hayward@cim.mcgill.ca

Abstract

There are many engineering challenges that must be addressed in order to successfully integrate haptics with the environmental characteristics found in surgery. The most fundamental of these challenges is to achieve update rates of solid nonlinear deformable objects that are acceptable to the human haptic system. This paper presents a software architecture that is designed to meet these challenges by analysing the task, the haptic, and the hardware constraints of surgery simulation.

1 Introduction

Imagine that a surgical procedure on a kidney is being simulated. A description of it shape, colour, and physical properties must be represented within the computer. Since the kidney occupies three dimensional space, the complexity and size of the description will be quite large. Nevertheless, for a human to successfully interact with the virtual kidney, this description must be processed at rates that are fast enough so that the simulation is perceived by the user as being realistic. A survey of current research in surgical simulation can be found in [3].

Probing deeper into the nature of the simulation exposes more difficulties. Fleshy regions are both inhomogeneous and nonlinear. Nonlinearities greatly increase the complexity of the solution and hence conflict with the realtime requirements of the haptic system. In particular, unlike linear systems, nonlinear systems do not always possess a unique solution and must be solved iteratively.

By analysing the tasks performed in surgery simulation, several observations can be made. First, for such tasks as cutting, a change in the physical structure of the body takes place. This precludes the use of simulation techniques that put the burden of the solution almost entirely into preprocessing. Secondly, because several computational tasks must be executed at the same time, surgery simulation must be multi-modal. For example, one region of the body may be required for its physical influence and appearance only, while another region may be part of a suturing or cutting simulation.

Consequently, to realise the surgical simulation of the kidney, the system requires that informed engineering tradeoffs be made. This can only be achieved by understanding the constraints imposed on the system. In the following Section, these constraints are discussed in detail in the context of finding the correct tradeoff balance.

2 The Problem Constraints

The system design is motivated by an analysis of the design constraints. They are are divided into haptic compatibility, processing constraints, properties of surgery simulation and force accuracy and visual acuity. By viewing these constraints from both a restriction and an opportunity perspective, a set of design guidelines will be created.

2.1 Haptic Compatibility

The purpose of a haptic display is to present forces and tactile stimulation to a human user. From an engineering standpoint, this implies that the haptic system must be designed with an understanding of the input–output properties of human beings: i.e., the system must be haptic compatible.

The notion of haptic transparency is pertinent to both haptic hardware and software. A perfectly transparent haptic system will present forces, impedances, and visual feedback that are indistinguishable from the real world situation it is representing. Intuitively, physics dictates that limitations on the haptic hardware alone will make ideal transparency impossible. Fortunately, the human user's perceptual system is imperfect. Limits in human dynamic force input range, tactile stimulation, and vision give the engineer the opportunity to design to psychophysical limits. Exceeding these limits only leads to over design because no improvement in system fidelity will be perceived.

Clearly, the haptic device itself provides the baseline for transparency: in particular, its dynamic range, its stiffness, damping, inertia, as well as limitations in the impedance, force, and acceleration that it can display. However, it is the task of the software to fully exploit the potential performance of the haptic device being used.

Haptic and telerobotic researchers have recommended several frequency ranges for system operation. Fischer *et. al.* [2] state that, although position inputs from the user can be read at 10Hz to 15Hz, the force output to the device should be executed at a minimum rate of 100Hz; ideally the rate should be above 500Hz. However, practise has shown that force update rates of 1kHz are more preferable. The visual system requires a different frequency range, generally in the order of 20-40Hz.

Force feedback latency is the time taken for a user's motion to elicit a force output to the haptic device. From a system's perspective, force latency creates additional phase lag in the closed loop system; this, in turn, reduces the phase margin and eventually causes instability. In tracking experiments, delays as low as 40ms can seriously degrade system performance [5]. Even if stability is maintained, haptic practitioners describe systems with too much lag as feeling soft and mushy [4]. In a deformable body haptic simulation, the largest delay is expected to be created by the computation of the deformable body itself; other contributing factors to force latency will be from communication, kinematics, control, sensing, and mechanical transmissions.

It is clear that the human user is a multi-frequency input–output system. In a system in which computational workload must inevitably be optimised, it neither makes sense, nor is it feasible, to process these demands at the highest minimum rate of 500Hz. This leads to the following observation.

Observation 1 *The deformable body haptic simulation should be a multi-frequency system; this will allow the efficient distribution of computational load.*

Another important observation is that these bandwidth constraints are not spatially uniform. Importantly, force feedback is needed only at the points of contact between the haptic device and the virtual body. In fact, forces on the periphery of this *region of interaction* are not even required. Therefore, to reduce force latency and to increase force feedback rates, it makes sense to process the region of interaction at a higher rate than parts of the simulation to the periphery. In these peripheral regions, only visual feedback is needed at the less demanding visual rates. Therefore the following observation can be made.

Observation 2 *Force feedback is required only at the points of contact between the haptic device and the virtual body. This region of interaction should be processed at the highest possible rates. On the periphery of this region, only the less demanding visual feedback rates are required.*

The human-in-loop haptic system presents difficult bandwidth challenges. However, it is also apparent that these demands are not the same throughout the simulation. This will provide a basis for the efficient distribution of computational resources.

2.2 Processing Constraints

A few hand calculations show that the problem size grows with the cube of the desired information density of the system. This cubic growth is a result of the volumetric bodies that are to be modelled. Most methods, such as finite elements and finite differences, track a finite number of discrete points in the system. This cubic increase also means that waiting for computer speeds to increase, or simply using many processors in parallel, is not a viable solution. Consequently, it is important to respect this cubic growth from the outset. This leads to the next observation.

Observation 3 *The system must reduce complexity at a cubic rate. Since the cubic increase is due to the volumetric simulation environment, it would be intuitive to reduce calculations at a volumetric (or cubic) rate too.*

In essence, any speed-ups or numerical complexity reduction must create a computational reduction of order n^3. Intuitively, quadratic or linear optimisations will rapidly be rendered ineffective by the growth of the volumetric system to be modelled. This cubic optimisation requirement is most likely to be found by going directly to the source of the the cubic growth itself – volume. This yields an important design criterion.

2.3 Properties of Surgery Simulation

The haptic system must enable the interaction of a haptic device with a virtual body. The virtual body may, for example, represent a brain, liver, kidney, or muscle. To achieve this, the simulator must have knowledge of the physical parameters of the components: for example, stiffness, damping, and mass density. The existing material data that is available, however, is written in terms of well known physical units: force, time, and displacement. This leads to the following observation.

Observation 4 *In order for the surgical simulator to be applicable to real world data, it should incorporate real world physical parameters.*

This observation discounts models that are not physically based, including many of the geometric based models found in computer graphics. This is also an important constraint because a simulator that uses its own unique set of physical descriptors may not be able to guarantee realism for arbitrary combinations of physical parameters.

The relationship between stress and strain in tissue material is also nonlinear. In general, linear relationships between stress and strain are valid only for small material strains and displacements. Beyond this region, nonlinear effects begin to dominate causing multiple solutions, path dependence, and iterative solution techniques. When this is considered in conjunction with the cubic processing requirements discussed in the previous section, the computational burden can appear overwhelming. However, the following observation affords a new design opportunity.

Observation 5 *Nonlinearities dominate in the regions of greatest strain, stress, or displacement. In a surgical simulation, this will generally be in the region of interaction because peripheral regions will be subject to less perturbation.*

Consequently, there is an opportunity to make a complexity–processing tradeoff by simulating the region of interaction using a nonlinear model and, to the periphery of this region, maintaining a linear model. This would greatly reduce the processing requirements caused by nonlinearities.

It is probable that a method for simulating the material dynamics would not be applicable to cutting, and vice-versa. This is similar to the previous idea that the simulation could be segmented into a linear and a nonlinear part. Therefore, the following observation is made.

Observation 6 *Surgery simulation is multi-modal. In one region of the simulation, a cutting or suturing task may be performed. In another part of the simulation, only the dynamics, or even statics, of the body may be required.*

This observation implies that several simulation computational engines should operate at the same time. They should be distinct yet work in unison.

Another important factor is adaptability. Some surgical simulation modes could create structural changes within the virtual body. This would occur, for example, in cutting, suture, and tissue removal simulations. Unfortunately, this has a negative impact on designs that place too much emphasis on preprocessing techniques to achieve realtime solution rates. This is because these implementations will not be capable of adapting to changing tissue structures in realtime. From a design perspective, this constraint proves difficult to work with. Often, however, surgical tasks, such as suturing, could often be simulated within localised regions only. Therefore, it is possible to soften this constraint for many tasks. This yields the following observation.

Observation 7 *The haptic surgical simulator should be able to accommodate localised changes in structure to the virtual body. This precludes methods that rely heavily on preprocessing for realtime speed up.*

Finally there is the issue of providing an open architecture. It is of paramount importance that this software architecture be open to future extensions. If it is not, then this line of research will meet a premature end. Therefore, the following observation can be made.

Observation 8 *The software architecture should be designed to allow for future extensions.*

In summary, the goal of surgical simulation imposes several difficult constraints: physical parameters, nonlinearities, and multiple modes. The multi-modal nature of surgery simulation requires that the system be open to a myriad of yet to be developed additions. Fortunately, it has been observed that nonlinearities are primarily required in a local region; this affords opportunities to optimise the system.

2.4 Force Accuracy and Visual Acuity

In order for the simulation to be realistic, it must be both fast enough and accurate enough to meet transparency requirements. Haptic accuracy demands are currently being established in the literature. However, it is hypothesised that force feedback direction may be as important as the force magnitude. As these constraints are established, it will be critical that the haptic simulator is able to meet them.

The simulation must also appear realistic. It is well known that visual acuity is not uniform across the human visual range. It is predicted that, when performing a task, the user's attention, and hence fovea, will be focussed on the region of interaction. In light of the preceding facts, the following observation can be made.

Observation 9 *Force accuracy and visual acuity are the most demanding in the region of interaction. On the periphery of this region, force calculations are not even required for a haptic interaction, and visual attention will be limited.*

This observation opens the possibility for a system design that varies accuracy spatially. In turn, this would allow computation–accuracy tradeoffs. For example, rather than processing the entire region to the accuracy demands of the haptic device, peripheral regions would be processed at lower tolerances; consequently, the peripheral region would require less processing time.

3 The Proposal

Finite element methods have been adopted as the computational engine of this work. They offer several properties that fit with the design criteria described in the previous section. Most importantly, the finite element method is physically based; in the case of solid mechanics, its equations are described in terms of stress, strain, and material parameters. An additional advantage is that finite elements allow for inhomogeneous, ansiotropic, and nonlinear materials. This meets the requirements described in Observations 4 and 5.

3.1 The Essential Idea

This work departs from traditional finite element techniques in that it adopts a new hierarchical (multilayer) mesh structure. The purpose of the hierarchical mesh is to distribute the computational workload optimally in order to realise realtime haptic simulation. At the same time, the hierarchical mesh incorporates the guidelines set out in the previous section.

The hierarchical mesh is most easily understood using a visual example. Figure 1 shows a regular finite element mesh being probed.

A recurring theme in Section 2 was that many of the problem constraints were not uniform across the entire simulation. In particular, force update rates and force latency (Observation 2), force accuracy and visual acuity (Observation 9) are most stringent in the region of interaction.

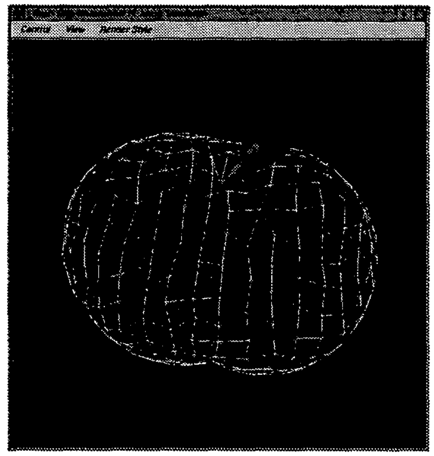

Figure 1: A 3-dimensional finite element mesh. The mesh is defined by tetrahedron elements. The virtual probe represents the position of the haptic device.

It was also noted in Observation 3 that computational reduction must be accomplished at a volumetric rate. Therefore, a volumetric region, in the neighbourhood of the haptic device, is cut out from the mesh. This remaining region is shown in Figure 2.

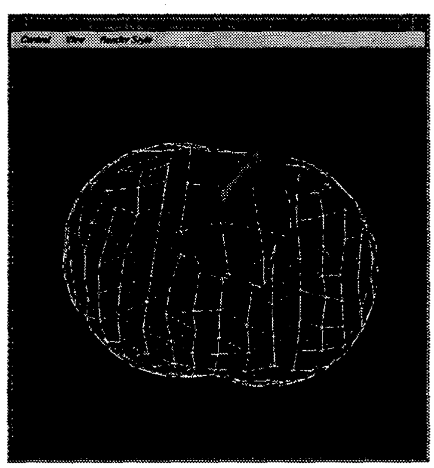

Figure 2: Cutting out the region of interaction. The most critical area with respect to haptic simulation is the region surrounding the contact point between the body and the device.

The remaining region shown in Figure 2 consists of the least time critical parts of the mesh. In this region force feedback is not needed, accuracy tolerances are lower, and only visual feedback rates are required. The region of interaction, cut out from Figure 2, is then defined as a separate finite element mesh. This mesh is known as a *child mesh*; it is shown in Figure 3.

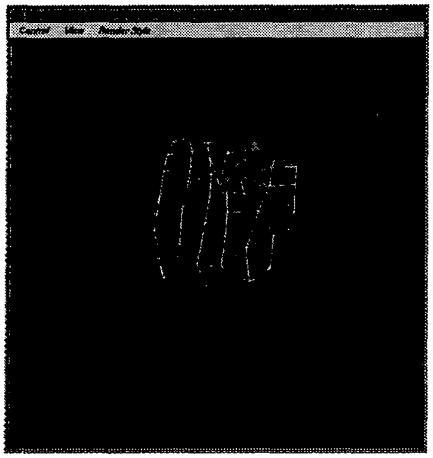

Figure 3: The child mesh. The region of interaction is modelled by a mesh that is a subset of the original mesh. This represents the most time critical region.

The child mesh, in Figure 3, models the region of interaction. This is the most haptic critical region of the simulation. The child mesh is defined as a self contained finite element system; consequently, it is computationally separate from its parent mesh in Figure 1. This computational independence is important because update rates, accuracy, and even the type of simulation can be defined independently for each of the child and parent meshes. This child mesh is also a subset of its parent; consequently, it is smaller and is faster to process. In many applications, the haptic device will explore, and interact with, several regions of the virtual body; therefore several overlapping child regions are defined.

The hierarchical mesh is designed so that the haptic device interacts directly with a child mesh, not the larger and slower parent. A perturbation on the child mesh will be processed as a displacement of one or more of the child mesh nodes. These node displacements represent the excitation to the child finite element system. This input is processed resulting in force feedback components that will be sent to the haptic device. The process also computes the displacements of all the nodes in the child system. Some of these displacements (those at nodes that exist in both the child and the parent) are then used as the input to the parent mesh; these inputs are then processed in the parent system in order to update its states.

The child meshes in the hierarchical system model are intended to model only a subregion of the body to be simulated. The unmodelled region must, however, influence the child mesh system so that the its physical properties and boundary conditions are present. This is achieved using an idea inspired by Thevenin and Norton equivalents in electronics. Here, the entire unmodelled system is described using an equivalent impedance network. This network exactly replicates the input–output properties of the network it replaces, but lacks the detail and computational requirements. Details on the equivalent impedance can be found in [1].

3.2 Meeting the Constraints and Observations

The hierarchical architecture respects the observations detailed previously in the following ways.

Ob.1 It was observed that a multi-frequency simulation was optimal. The parent and child meshes are described by distinct finite element systems. Consequently, at runtime, they can be executed at different frequencies. This enables the hierarchical mesh to distribute processing requirements based on the human input frequency demands; in particular, the leaner child mesh can execute at a faster rate than its parent. In turn, this increases the maximum force update rate and decreases the force latency.

Ob.2 It was observed that force fidelity is needed in the region of interaction only. This region of interaction corresponds to the child mesh. Therefore, force fidelity can be maximised at the child, without simultaneously increasing processing demands on the parent mesh.

Ob.3 It was observed that computational reduction must be achieved at a volumetric rate. The hierarchical mesh achieves this by defining smaller child meshes that are *sub-volumes* of the parent. Computational reduction is then achieved by processing only this smaller child mesh at the more demanding haptic rates, while the parent can be processed at a slower rate.

Ob.4 Using the finite element technique allows the system to be described in terms of physical parameters. Yet, at the same time, the hierarchical mesh does not change any finite element theory or concepts. This is because the equivalent impedances are calculated once the finite element preprocessing is completed.

Ob.5 It was noted that tissue reacts nonlinearly to perturbations; however, these nonlinearities will generally dominate at the region of interaction only. The hierarchical mesh partitions the system into a large linear system (the parent) and a smaller nonlinear system (the child).

Ob.6 It was observed that surgery simulation must be multi-modal. Because the child meshes and the parent mesh are computationally separate, it is possible to have the child mesh executing a dynamic simulation and the parent mesh executing a quasi-static simulation simultaneously.

Ob.7 Local structural changes can be made at the child level. It is feasible to make these changes at the parent level too; however, this would require processing on an additional processor.

Ob.8 It was observed that the software architecture should be open to future extensions. It is difficult to design for this when the details of the future extensions are unknown. Nevertheless, in an attempt to achieve extendibility, the hierarchical mesh does offer both the structure and the equivalent impedance network. For example, a computational method for suturing could be developed without using the finite element technique. If this method can incorporate the equivalent impedances, then there is no reason that the new technique cannot be used in the hierarchical mesh system. Further, this could potentially allow for hybrid simulators that bridge the gap between geometrical and physical techniques.

Ob.9 Force accuracy can be optimised at the child mesh. This is where force feedback is needed. Meanwhile, on the periphery, the parent mesh does not require processing to the same accuracy, and hence a computational optimisation is achieved. (This factor was an important issue in ideas that preceded the hierarchical mesh because systems that produce a set of simultaneous equations must be solved, by their very nature, simultaneously. Therefore, it is not possible to specify that certain variables in the equations be resolved to different accuracies than others. The partitioning of the system into computational separate numerical systems alleviates this problem.)

4 Conclusions

This paper described the surgical simulation problem and defined the constraints it imposes on the designer. This led to a set of design observations that inspired the hierarchical finite element mesh. On a single processor SGI IMPACT R10000, a linear system containing 431 nodes achieved child update rates between 40 and 100 Hz; simultaneously the parent system executed at rates between 1 to 10Hz. The frequency used depends on both the desired accuracy and child mesh size. A nonlinear child system has also been implemented with an update rate of 20Hz. Future papers will present these results in more detail.

5 Acknowledgements

This research is funded by project "Reality-based Modeling and Simulation of Physical Systems in Virtual Environments" supported by IRIS-III, the Institute for Robotics and Intelligent Systems part of Canada's National Centres of Excellence program (NCE). Additional funding is provided by NSERC, the Natural Science and Engineering Council of Canada, in the form of a postgraduate scholarship for the first author and an operating grant for the second. The first author also received funding from a Raychem Cooperation Paul M. Cook International Scholarship.

References

[1] O. Astley and V. Hayward. Multirate haptic simulation achieved by coupling finite element meshes through norton equivalents. *IEEE International Conference on Robotics and Automation*, May 1998.

[2] P. Fischer, R. Daniel, and K. Siva. Specification and design of input devices for teleoperation. *IEEE International Conference on Robotics and Automation*, 1:540–545, 1990.

[3] R. D. Howe, Y. Matsuoka, Y. Robotics for Surgery. *Annu. Rev. Biomed. Eng. 1999*. 01:211-240.

[4] B. Jackson and L. Rosenberg. Force feedback and medical simulation. In R. Satava, K. Morgan, H. Sieburg, R. Mattheus, and J. Christensen, editors, *Interactive Technolgy and the New Paradigm for Healthcare*, Studies in Health Technology and Informatics, chapter 24. IOS Press, 1995.

[5] C. Wickens. The effects of control dynamics on performance. In K. Boff, K. Kaufman, and J. Thomas, editors, *Handbook on perception and human performance – cognitive processes and performance*, volume 2, chapter 39. Wiley-Interscience, 1996.