Design Strategies for Real-Time High-Performance Multimedia Applications on Multiprocessors

First Year Report

March 7, 2000

Per Stenström and Jan Jonsson
Department of Computer Engineering
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
{pers,janjo}@ce.chalmers.se

Ulf Assarsson
ABB Robotics
SE-412 50 Gothenburg, Sweden
ulf.assarsson@se.abb.com

1. Summary of Project

According to the original project proposal, the goal of this project is to develop concepts in terms of (i) *quality-of-service negotiating scheduling algorithms* that adjust the quality-of-service (QoS) level dynamically and (ii) *worst-case performance analysis and improvement techniques* to reduce the gap between average and worst-case performance for virtual-reality applications on multiprocessor systems. The developed concepts shall make it possible to design parallel programs that can achieve high performance on a wide range of multiprocessors under real-time requirements.

The project goals will be achieved through two Ph.D.student projects carried out by Jonas Lext and Björn Andersson whose detailed plans are available from http://www.ce.chalmers.se/pamp/utv/jonas.pdf and http://www.ce.chalmers.se/pamp/utv/bjorn.pdf, respectively. The focus of the first Ph. D. subproject is real-time rendering algorithms and their efficient implementations on multiprocessors. Key research issues are to find efficient parallelization strategies to address the performance and real-time goals of real-time rendering algorithms.The focus of the second Ph. D. subproject is to develop principles for QoS negotiation in adaptive real-time multiprocessor systems. Here, we want to devise run-time scheduling heuristics that can adapt their behavior dynamically in accordance with available resource capacity and designer specifications.

Apart from the two Ph. D. students directly financed by ARTES/PAMP, the project also involves an industrial Ph. D. student (Ulf Assarsson) whose project focus is on real-time rendering algorithms for multiprocessor systems. In addition, apart from the two senior researchers Per Stenström and Jan Jonsson, a third senior scientist, Tomas Möller, is associated with the project. He acts as co-supervisor for Jonas Lext and Ulf Assarsson and is currently spending a one-year post-doc at UC Berkeley in Professor Carlo Sequin's group (http//www.cs.berkeley.edu/~sequin).

Ulf Assarsson's involvement in this project also constitutes the industrial link to ABB Robotics who uses the technology developed in the project in a CAD application taking the form as tools for real-time 3D simulation of industrial robots in automated factories.

In this first-year project report, we briefly overview the original plans, the achieved results, and the deviation from the original plan for each of the two projects financed by ARTES/PAMP in Sections 2 and 3, respectively. Finally, we comment on the industrial role of the projects in relation to the collaboration with ABB Robotics in Section 4.

2. Project status "Parallelization of VR algorithms"

2.1 Original plan

In the original plan, the following milestones and deliverables were specified:

- **Activity 9807-9809:** ARTES/PAMP student recruitment.
- **Activity 9809-9909:** First phase of project. In this phase the application requirements will be analyzed by selecting study objects in terms of parallel visualization algorithms to identify performance and real-time issues. **Deliverable 9909:** Selection of study objects in terms of parallel implementations of visualization algorithms. State-of-the-art report on performance and real-time issues based on preliminary evaluations. Refined specifications of the Licentiate project topics and project plans.

The project formally started in January 1999 because the student recruitment took a little longer than expected. In August 1999, a first revision of the original project plan was done motivated from an evaluation of all ARTES projects that took place during the ARTES/PAMP summer school in Linköping in August 1999. In the revised study plan, the following concrete project goals were specified:

- ***Study 1: Design and compilation of a benchmark for dynamic scenes***

This study aims at identifying suitable performance and real-time metrics to evaluate real-time rendering algorithms. The key result of this study is a methodology to evaluate properties of real-time rendering algorithms along with a benchmark suite that allows for objective comparisons between different algorithms. The anticipated deadline for completion of this project phase was 9912.

- ***Study 2: Evaluation of new acceleration methods for real-time ray tracing***

This study aims at evaluating two new ideas regarding the intersection tests in ray tracing that often is a performance bottleneck in this application. The intersection tests to be evaluated are not only expected to perform better than previously proposed methods but are also expected to lend themselves better for animation in that they take advantage of scene geometries and frame-to-frame coherency to speed up the computation. Thus, they are good candidates for meeting the real-time constraints for rendering in real-time. The evaluation will be based on the methodology developed in study 1. The anticipated deadline for the completion of this project phase is 0009.

2.2 Project results and deviation from original plan

Study 1 is almost completed according to the plan. The results from this project phase are reported in a paper entitled *"BART: Benchmark for Animated Ray Tracing."* This paper will soon be submitted to IEEE Computer Graphics and Applications. This paper proposes a benchmark suite embodied in a suite of test scenes and a procedure for reporting performance results. In the process, a number of reasons for performance problems for rendering algorithms are identified which has prompted the definition of the benchmark suite. The hope is that this suite and the procedures to report performance will be widely used so that performance of interactive rendering algorithms can be objectively compared.

Study 2 will be inititiated during the spring and we anticipate that a state-of-the-art article will be produced and submitted during the summer. A third study will be concerned with parallelization of a real-time ray tracing algorithm. The key goal of this study is to isolate the performance and real-time issues on several multiprocessor platforms including Sun Enterprise servers and SGI Origin 2000. These three studies will constitute the licentiate dissertation of Jonas Lext.

2.3 Original plan

In the original plan, the following milestones and deliverables were specified:

- **Activity 9807-9809:** ARTES/PAMP student recruitment.
- **Activity 9809-9909:** First phase of project. In this phase the different task dispatching methods will be analysed and their performance assessed. **Deliverable 9909:** State-of-the-art report on properties and other issues based on preliminary evaluations. Refined specifications of the Licentiate project topics and project plans.

The project formally started in January 1999 because the student recruitment took a little longer than expected. In August 1999, a first revision of the original project plan was done motivated from an evaluation of all ARTES projects that took place during the ARTES/PAMP summer school in Linköping in August 1999. In the revised study plan, the following concrete project goal was specified:

- ***Study 1: Analysis of real-time multiprocessor task dispatching methods***

This study aims at analysing the properties and performance of two different task dispatching methods for shared-memory multiprocessor systems, namely the partitioned approach (where each task always execute on the same processor) and the non-partitioned approach (where a tasks can execute on an arbitrary processor). The purpose of the study is to determine which of these methods are most suitable for an adaptive real-time system where task sets may change during the course of the system's mission. The key results of this study are (i) an in-depth understanding of mechanisms of the different dispatching methods including identification of possible scheduling anomalies, and (ii) how the temporal behaviour of the methods can be predicted in advance. The anticipated deadline for completion of this project phase was 9908.

- ***Study 2: Analysis of bus contention on schedulability for shared-memory multiprocessors***

This study aims at analysing the impact of bus operations in a shared-memory multiprocessor on the schedulability of a set of real-time tasks. The main focus of the study is to analyse whether it is possible to determine a worst-case scenario for memory references in a shared-memory architecture with a round-robin-based bus arbitration policy. Since memory references typically are a function of local cache contents, all bus operations do not occur in reality. We intend to analyse and evaluate the overhead as caused by bus operations for the task dispatching methods in Study 1. The anticipated deadline for completion of this project phase was 0006.

2.4 Project results and deviation from original plan

Study 1 has been almost completed. The results from this project phase has been reported in a paper entitled "Fixed-Priority Preemptive Scheduling: To Partition or Not to Partition". This paper was submitted to the *Int'l Conference on Parallel Processing* in January 2000. The main results in that paper are all aimed at refuting the common belief that the non-partitioned method is inferior to the partitioned method for real-time scheduling. To this end, we have made the following contributions. (1) We have identified important scheduling anomalies and limitations for the non-partitioned method and pinpointed fundamental differences between the schedulability conditions in multi- and uniprocessor systems. (2) We have proposed a new priority-assignment policy for the non-partitioned method that circumvents many of the identified anomalies and limitations; and, by simulation experiments, we have shown that the new policy increases the system's ability to meet deadlines to levels above that of the partitioning method. (3) We have proposed a new task dispatching algorithm for the non-partitioned method that reduces the number of task preemptions to levels below that of the partitioning method.

Except from being delayed a few months, this study is following the plan and no major deviation from the original plans has occurred. However, some additional results from this study remain to be demonstrated. More specifically, we are currently working on formal derivation and proofs for some of the hypotheses stated in the earlier parts of the study. These results will be reported in a paper entitled "Properties, Priority Assignment and Utilization Bounds for Non-Partitioned Fixed-Priority Scheduling". We plan to submit this paper to the *IEEE Real-Time Systems Symposium* in May 2000.

Study 2 has not yet been inititiated, but we plan to decide during the summer of 2000 whether this study will indeed be pursued. As an alternative, we may instead initiate a third study which will encompass an implementation of the dispatching and priority-assignment mechanisms proposed in Study 1. The reason for this is that we want to assess the impact of software and architecture properties on the performance of the proposed mechanisms, using a multiprocessor system running the Linux operating system. One of these studies (2 or 3) together with Study 1 will constitute the licentiate dissertation of Björn Andersson.

3. Statement from participating industry

ABB Robotics a leading supplier of industrial robots with more than 70,000 robots installed world-wide. The company with 600 employees is part of the global ABB Flexible Automation organization. ABB Robotics develops and manufactures robots. An essential part of the robot is the software, where efficient methods to support the programming of the robot is key. One way to improve the programming aspect is to use VR models. Programming is then performed by e.g. pointing out the route of the robot in a virtual reality model of the factory. Because many programming tasks, such as pointing out welding points, involve very detailed models of the factory, a key challenge for such tools is the availability of high-resolution real-time VR models that can be run on cost-effective platforms.

The focus on real-time rendering algorithms in this project is clearly an important asset to the technology used in the tools developed at ABB Robotics. Also, the focus on investigation of design principles of such algorithms on multiprocessors is interesting owing to the fact that multiprocessors are likely to become the de-facto platform even for low-end PC machines.

Our industrial Ph. D. student Ulf Assarsson who is the contact person in this project for transferring know-how is currently working on how to utilize multiprocessors for common graphics algorithms involving tree traversal, in particular "view frustum culling" - i.e. finding the potentially visible objects in a particular view - and collision detection (by tradition a subject in computer graphics). The resulting algorithms will be directly merged into the applications. This will be followed by a study of how to speed up ray-tracing with a method called "reprojection". Ray-tracing is needed when demands for image quality is very high and when correct reflections and shadows are important, for instance when planning and designing the shopfloor regarding light positions.