Modeling of medical visualization in heterogeneous node architectures.

Internship description – starting around September 2017
Rob de Jong, Philips Health Care (rob.de.jong@philips.com)
Marc Geilen TU/e (m.c.w.geilen@tue.nl)

This work is part of a Philips cooperation with several partners in Europe, amongst others universities from Milan and Valencia.

Multiple medical scanners (CT, MRI, Ultrasound, x-ray, nuclear medicine) produce a stack of 2D images representing a 3D volume. To visualize such a volume, as a projection on a 2D image seen from an arbitrary angle, a so-called ray-casting algorithm is used. Our existing ray-casting algorithm is written in C and originally runs in a Window environment (Visual Studio 2013).

This internship aims to develop a model. Based on that model adapt the existing algorithm for further acceleration on a multi-processor machine with a variety of processing nodes. These processing nodes can be CPUs, GPUs and functions implemented in dedicated hardware (FPGAs).

Work already done:

This internship will build upon earlier work that focused on isolating our ray-caster from the production code, porting from Windows to Linux, and partition this code in control code and kernel code. The kernel code works on blocks of the screen and is very well suited for parallelization. The control code is relative lightweight and runs on a CPU which drives the compute intensive kernel code that can be accelerated to run on an array of dedicated nodes (CPUs or GPUs). The scheduling of the kernels to the nodes and the communication between the different nodes is provided within this project by one of our partners.

Specific activities:

The goal of this assignment is to create a model and determine its optimal parameters for distributing the kernel instances at the different node types that are available, aimed at reducing the latency. E.g. one of such parameters could be about the block size in relation to the cache sizes of different processing nodes. The existing code, although prepared for parallel processing, runs on a single machine and the proposed solution should be highly distributed. This means an effective and efficient distribution amongst the available heterogeneous nodes needs to be chosen. As the algorithm is memory bound, special care needs to be taken to distribute the input data amongst the nodes. A stack of images that make a 3D volume can be up to 1.5 GB.

Other parameters are amongst: bandwidth of the various interconnects, available FPGA area, size and availability of local cache memory.

In the European project MANGO a dynamic distribution of node resource is developed. This means, the algorithm needs to react in real time to changes in system composition and load. The model should be able to adapt in real time.

Part of the assignment is to implement the model in the MANGO prototype and measure its performance.

Prerequisites

- Affinity with bottleneck analysis of a network of heterogeneous processing nodes and its interconnects.
- Modelling capabilities of such a network
- Ability to write C / C++ code
- Pre: Knowledge of Linux / gcc toolchain.