

Purpose: To develop a highly data-parallel version of B-spline registration within the stream-processing model, suitable for use on cheap stream processors such as graphics processing units (GPUs).

Method and Materials: We first develop a grid-alignment technique and a data structure that greatly improves the computation speed of B-spline registration on the CPU. The basic idea here is to align the B-spline grid with the voxel grid, so that the volume is partitioned into tiles of equal size. The use of equal sized tiles is important since it allows the coefficient multipliers used for B-spline interpolation to be pre-computed. Then, using this data structure, we develop a data-parallel version of B-spline registration suitable for use on a GPU. We have developed a streaming algorithm for B-spline registration within the Compute Unified Device Architecture (or CUDA), a programming abstraction for general-purpose computing on GPUs, recently introduced by NVidia.

Results: Using both phantom (swine lung) and synthetic images, we have validated and compared the GPU-based algorithm with the B-spline registration routines in the Insight Toolkit (ITK version 3.10.1). Our CPU implementation that uses the grid-alignment technique runs up to 400 times faster than ITK. The GPU-based algorithm achieves further speedup; it runs about four times faster per iteration than our CPU version, that is, about 1600 times faster than ITK B-splines. The registration quality achieved by the GPU was found to be excellent, with nearly identical displacement vectors compared to the CPU version.

Conclusions: Improvements in the processing speed of deformable registration algorithms are a powerful driving force for clinical applications. Interventional procedures such as image-guided surgery and image-guided radiotherapy require very low latencies in imaging and analysis, and improvements in processing speed mean that deformable registration can be used in such settings.