Stanford 50 – Graduate Student Poster abstract –

Mr. Larsen's expected date of Ph. D. completion: June, 2011
Dr. Elster and Mr. Fevang are Mr. Larsen's Master supervisors.
Mr. Larsen is the main contributor to this work.

Speeding Up Transform Algorithms for Image Compression Using GPUs

Leif Christian Larsen, (1) Anne Cathrine Elster(1) and Tore Fevang (2)

(1) Dept. of Computer & Information Science (IDI).
Norwegian University of Science and Technology (NTNU).
leifchl@idi.ntnu.no and elster@idi.ntnu.no
(2) Schlumberger Limited, Trondheim, Norway
TFevang@trondheim.oilfield.slb.com

Poster abstract

Graphics Processing Units (GPUs) have traditionally been used for computer games and visualization applications. However, since they recently also have added features for general purpose programming, they are now an interesting platform for general scientific computing due to their low cost, high memory bandwidth and high floating-point computational power, they have lately been used for other computationally intensive tasks.

On this poster, we show how GPUs can be favorably used to off-load computations in applications utilizing compression methods using transform-based algorithms. GPUs are well suited to such computations, due to their highly parallel architecture. GPU implementations of parallel Discrete Cosine Transform (DCT) and Lapped Orthogonal Transform (LOT) algorithms are compared to corresponding both dual and quad core CPU implementations.

For large data set sizes and 32-bit floating-point precision, we show that DCT on the GPU outperforms DCT on single-core CPUs with wall-clock and CPU time speedups 1.18 and 1.22, respectively. In addition, we show a wall-clock speedup of 2.77 and a total CPU-time savings of 11.14 over our quad-core CPU LOT implementation for our GPU LOT implementation.

Our results show that off-loading computationally intensive transforms applied to large data sets to the GPU gives both wall-clock speedups and CPU time savings. The GPU should therefore be well suited for transforms of even greater arithmetic complexity, such as the theoretically optimal Karhünen-Loève transform as well as the Generalized Lapped Orthogonal Transform (GenLOT) which has superior S/N (signal-to-noise ratio) than the DCT and LOT algorithms. Potential improvements of these algorithms with respect to GPU implementations will also be discussed.