Single-Loop Software Architecture for JPEG 2000

David Barina, Ondrej Klima, and Pavel Zemcik

Brno University of Technology, Brno, Czech Republic
{ibarina,iklima,zemcik}@fit.vutbr.cz

JPEG 2000 is an image coding system based on the discrete wavelet transform. Unfortunately, there exist several major issues with the effective implementation of the codec. For high resolution data decomposed by a separable transform, immensely many CPU cache misses occur. Following the procedure as defined in the standard, the coefficients of a single resolution appears all at once. Consequently, the entropy coder (EBCOT) needs to once again return to the data already touched.

We present a software architecture designed for JPEG 2000 coders. The proposed method employs a strip-based data processing technique while it performs a single-pass multi-scale wavelet transform. The overall compression chain is driven by incoming data while the fragments of the resulting bitstream can be produced immediately after loading the corresponding data and additionally in parallel. The method is friendly to the CPU cache and can nicely exploit the SIMD extensions.

The established strip-based transform directly produces the code-blocks one by one. The input is consumed gradually using strips with the height of 2×2^{c_n}, where 2^{c_n} is the code-block height. No more input data are required to be placed in physical memory. In total, our solution requires $(2S + 3 \times 2^{c_n})M$ samples populated into the physical memory, where M is the width of the image, and S is the number of lifting steps. The processing of code-blocks is then chained together to progressively produce the multi-scale transform. On any level, the processing can be further parallelized in such a manner that the code-blocks are generated in parallel. As a consequence, this parallelism involves interleaving of the DWT and Tier-1 stages. When the transform of the subsequent code-block is not started earlier than the EBCOT on the current one has been finished, the memory for HL, LH, and HH sub-bands can be reused.

The decomposition is performed using the 2×2 core with a lag $F = 3$ samples (for CDF 9/7 wavelet) in both directions. The core consumes a 2×2 fragment of the input signal and immediately produces a four-tuple of coefficients (LL, HL, LH, HH).

We have integrated our solution into OpenJPEG library (the reference JPEG 2000 software). The performance of the implementation outperforms the original code even if no parallelization and no SIMD extensions are used. See Figure 1. When the parallel processing is enabled, the performance increases proportionally to the input size and number of processing threads.

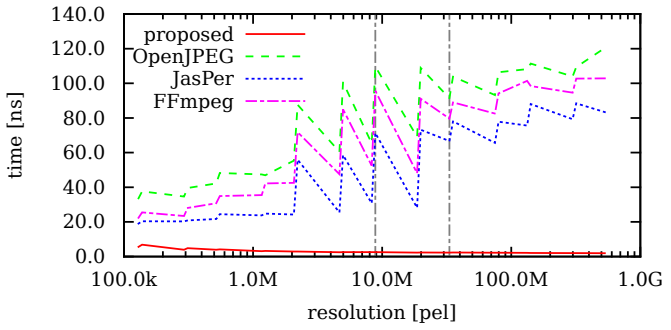

Figure 1: Performance of the transform stage only. DCI 4K and 8K UHD indicated by the vertical lines.