Master thesis review

Author of the thesis: **Jan Kovanda**

Thesis title: **Optimization of Light Propagation Computation**

Thesis content

This thesis deals with the optimization of the light propagation process between two parallel planes under limited memory conditions. Simulation of light propagation can easily require several terabytes of memory space, making the problem directly unsolvable on current ordinary computers. The proposed solution consists on decomposing the source and target planes into tiles and independently computing light propagation among individual tiles, later compositing the contributions. Since light propagation is done as a cyclic convolution the Fast Fourier Transform (FFT) is used and this is the most time consuming operation of the whole process. The specific problem being addressed in this Master thesis is determining the sizes of the tiles (number and shape), such that good performance is achieved in spite of the tiling of the planes.

In order to determine these sizes an extensive empirical study of available FFT libraries is performed, efforts focusing in the „Fastest Fourier Transform in the West“ (FFTW) library, which is demonstrated in chapter 2 to be the most adequate. This extensive study results on a number of empirically obtained recommendations, including: fftw-friendly dimensions, ratio between row and column dimensions (based on the „bathtub“ profile, but notice that results exhibit large variance), relevance and impact of the planner flag, impact of multithreading (with multicore processors) in execution time, etc.

From the above recommendations an optimization planner is built. This planner first partitions the source and target planes such that the data structures required for each FFT fit into memory and then keeps on subdividing such that FFTW friendly sizes are achieved. Unfortunately, the results presented in the document do not go far enough in demonstrating the quality of the performance results obtained.

Finally, in chapter 5 the main findings are summarized.

Overall Assessment

This Master thesis performs a very extensive empirical study of FFTW behaviour and, from this expoerimental data, derives some recommendations (and develops an optimizer to apply them) to the particular case of light propagation. This is a very interesting and complete work, which can be leveraged by the scientific community.

There are two major reasons why my final grade recommendation is 2 (very good), instead of excellent:

1) All conclusions and recommendations are based on empirical data, obtained with a particular version of a particular library running on a particular set of machines. Some of the recommendations are only loosely supported by the results, because these presente large variations (e.g., the ratio between the number of rows and columns). The very empirical nature of the data makes me wonder

how generalizable the conclusions are to other settings, in particular, whether these recommendations can still be used in future generations of both the hardware and the software. It is this lack of generalization that worries me a bit; on the other hand, the methodology could certainly be used, meaning that repeating the experiments on more recent versions of the environment could allow for verification of these recommendations;

2) in spite of what I say in the last sentence of the previous item, the used methodology is not always very clear. See my comments below regarding chapter 2. Also chapter 3 is sometimes a bit confusing, with many different options being used to draw the charts (making them more difficult to read). I recommend, if this work is to be published somewhere, that an effort is done to make the whole results' presentation more uniform.

Details

Chapter 1 clearly explains and formulates the light propagation problem and in particular its huge memory requirements. Equation 2 and the explanation that follows up to figure 1.7 clearly illustrate the problem at hand. I think chapter 1 would benefit from an additional equation equivalent to equation 2, but for the specific case of the tiled planes approach: for instance, is the IDFT applied to each tile in the „replaced content" buffer (I am using the terminology of figure 1.6) individually or to the whole buffer? It must be the former since the latter would rise memory issues again! And is the IDFT applied after the contributions of all the tiles in the source for a given tile in the target have been computed or before? Even though this can be easily inferred from the process it should be made clear in this chapter and presented as mathematical expressions.

Chapter 2 assesses different FFT libraries. It starts by presenting a series of requirements that all candidate libraries should possess and then assesses a large number of libraries. A publicly available benchmark is used to report all results. This is a very good practice since others will be able to reproduce and/or extend these results. Unfortunately, this benchmark reports throughput (mflops) instead of execution time and it is this last metric (performnace) that we are really interested. Throughput is converted into time by using equation 4. This equation uses a theoretical number of operations in order to do the conversion. This apporach is, in my opinion, not the best or clearer, for two reasons. First, since the theoretical number of operations includes a logarithm conversion between the two metrics is not linear and all the figures (which present mflops) become much harder to read (an improvement of 2 in throughput for different sizes N does not translate into an improvement of 2 in execution time, as stated by the student see eqs (5) and (6)). Furthermore, since the theoretical number of operations does not necessarily reports the real number of operations performed (it is an estimation), conclusions based on execution time might be misleading if comparing different implementations (which is the case) or execution in different computers. So, even if I agree with the final conclusion that FFTW is the most promising library I still think that the experimental methodology could have been improved: ideally, the benchmark could have been instrumented to directly measure execution time.

Finally, some figures present the dimensions of the 2D matrix (e.g., 256x256), whereas others present the total number of points (e.g., fig 2.4). I understand this is due to limitations of the particular tool used (Excel I believe), but once again it makes results harder to compare. Usually, holographers are experienced users of MatLab... this tool would have allowed for much more coherent figures.

Chapter 4 presents the optimization planner. The included results are not extensive enough to fully convince the reader of the performance gain associated with this. The student later sent me an exhaustive comparison of using versus non-using the optimization planner which clearly shows the performance gain associated with it. In particular a 10000x10000 source and a 1000x1000 target are used. The FFT performance with and without partitioning into fftw friendly tiles is compared; the former results in the evaluation of 40 FFTs, wheras the latter requires a single one. Execution times show the advantage of using the optinmization planeer. In my opinion, this Master thesis would gain significantly from including these results in the final document.

I propose the grade **2** and recommend the thesis for the defense.

Braga, Portugal, 5 September 2013

Luís Paulo Peixoto dos Santos, Ph.D

Luís Paulo Peixoto dos Santos