

Tutorial 9

In this tutorial we will implement a CUDA program that implements a simple local mean filter for a 1D array of arbitrary length.

- 0.) Download the [skeleton code](#) and generate the build system using `cmake`.
- 1.) Implement the steps required on the host to run a Cuda program:
 - i.) Query the number of Cuda devices, check the required properties, and initialize a suitable one.
 - ii.) Allocate memory on the device and copy the input data to the device.
 - iii.) Set up the number of thread blocks and threads that are required and call the device program.
 - iv.) Copy the result back to the host and clean up device memory.
- 2.) Implement the kernel program:
 - i.) Allocate the required shared memory to perform the mean computation in fast shared memory.
 - ii.) Load the input data for the thread block into shared memory.
 - iii.) Implement the mean computation.
 - iv.) Write the result back to global memory.

Please finish the implementation until next week (week of 02/01/2020).