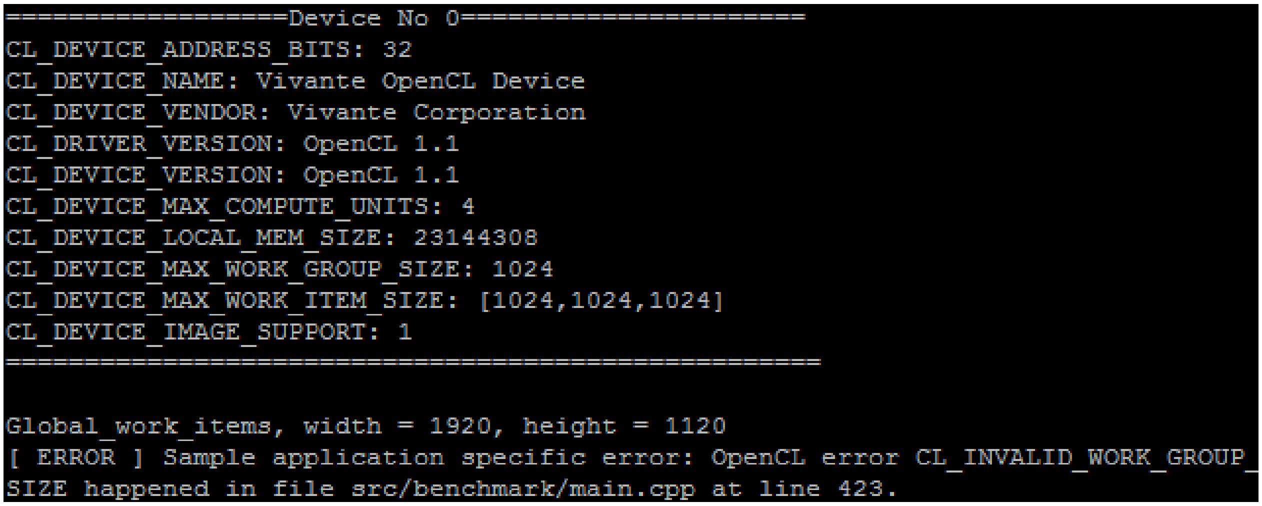
OpenCL Benchmark Progress Report (2016.09.30)

1. A potential reason for why kernel executed slowly on Vivante platform: the local_work_size parameter in clEnqueueNDRangeKernel(). By default, it set to be NULL, which means the OpenCL implementation will determine how to be break the global work-items into appropriate work-group instances. I try to test different combination on my laptop Nvidia GPU, the result as following (under buffer&&int condition):
    1) local_work_size[2] = { 1, 1 }; time(for 100 filter loop) = 11252 ms
    2) local_work_size[2] = { 32, 32 }; time(for 100 filter loop) = 187 ms

    When I test on the target, the result of the default local_work_size is almost the same as the {1, 1} case. But when I test other case, like {32,32}, it always shows error as following:



```
==================Device No 0======================
CL_DEVICE_ADDRESS_BITS: 32
CL_DEVICE_NAME: Vivante OpenCL Device
CL_DEVICE_VENDOR: Vivante Corporation
CL_DRIVER_VERSION: OpenCL 1.1
CL_DEVICE_VERSION: OpenCL 1.1
CL_DEVICE_MAX_COMPUTE_UNITS: 4
CL_DEVICE_LOCAL_MEM_SIZE: 23144308
CL_DEVICE_MAX_WORK_GROUP_SIZE: 1024
CL_DEVICE_MAX_WORK_ITEM_SIZE: [1024,1024,1024]
CL_DEVICE_IMAGE_SUPPORT: 1
====================================================

Global_work_items, width = 1920, height = 1120
[ ERROR ] Sample application specific error: OpenCL error CL_INVALID_WORK_GROUP_
SIZE happened in file src/benchmark/main.cpp at line 423.
```

    I have checked the requirement of local_work_size:

    1)The total number of work-items in the work-group must be less than or equal to the CL_DEVICE_MAX_WORK_GROUP_SIZE value;

    2) (global_size are {1920,1120}) global_work_size[0],... global_work_size[work_dim - 1] must be evenly divisable by the corresponding values specified in local_work_size[0],... local_work_size[work_dim - 1];

    3) local_work_size[0],... local_work_size[work_dim - 1] must be less than or equal to the corresponding values specified by CL_DEVICE_MAX_WORK_ITEM_SIZES[0],.... CL_DEVICE_MAX_WORK_ITEM_SIZES[work_dim - 1].

    I didn’t find the reason for this error.

Conclusion: the default divided local_size may not be optimal; next step, we should find a way to define the local_work_size[2] of the target.

2. When the local_work_size is {1,1}, I found an interesting fact:
   Detail: for instance, under buffer&&int condition, I compared two situations:

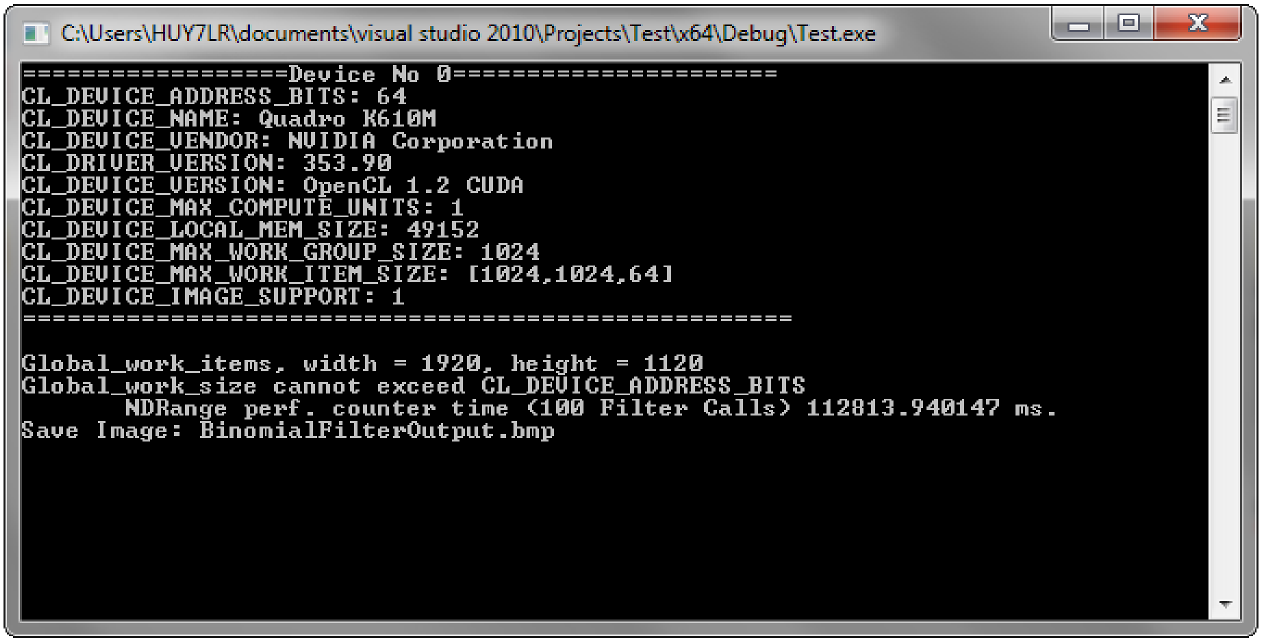
   1) put the clFinish() inside the kernels loop, which means the second kernel can only be enqueued until the first one is finished.(one-by-one);

   2)put the clFinish() outside the loop, which means there may be many kernel executing command in the queue, the GPU is supposed to service the next waiting command once there are enough work_items available.(may be parallel)

```
// execute kernel
for (int l_i = 0; l_i < 100 ; l_i++)
{
    clStatus = clEnqueueNDRangeKernel(command_queue, kernel, 2, offset, global_work_size, NULL, 0, NULL, NULL);
    SAMPLE_CHECK_ERRORS(clStatus);
    clStatus = clFinish(command_queue);
}
```

I try to execute on the laptop Nvidia GPU under the 2 situation.

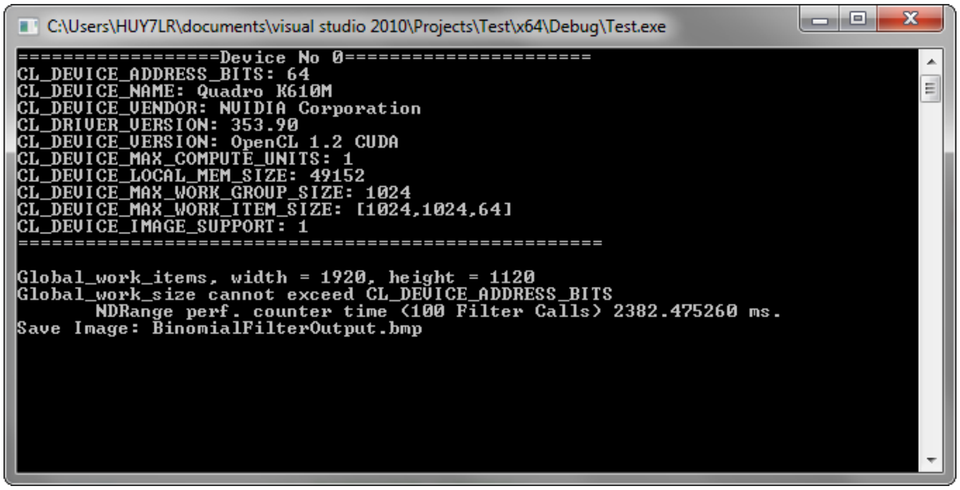
1)


C:\Users\HUY7LR\documents\visual studio 2010\Projects\Test\x64\Debug\Test.exe

```
==================Device No 0======================
CL_DEVICE_ADDRESS_BITS: 64
CL_DEVICE_NAME: Quadro K610M
CL_DEVICE_VENDOR: NVIDIA Corporation
CL_DRIVER_VERSION: 353.90
CL_DEVICE_VERSION: OpenCL 1.2 CUDA
CL_DEVICE_MAX_COMPUTE_UNITS: 1
CL_DEVICE_LOCAL_MEM_SIZE: 49152
CL_DEVICE_MAX_WORK_GROUP_SIZE: 1024
CL_DEVICE_MAX_WORK_ITEM_SIZE: [1024,1024,64]
CL_DEVICE_IMAGE_SUPPORT: 1
====================================================

Global_work_items, width = 1920, height = 1120
Global_work_size cannot exceed CL_DEVICE_ADDRESS_BITS
       NDRange perf. counter time (100 Filter Calls) 112813.940147 ms.
Save Image: BinomialFilterOutput.bmp
```

2)


C:\Users\HUY7LR\documents\visual studio 2010\Projects\Test\x64\Debug\Test.exe

```
==================Device No 0======================
CL_DEVICE_ADDRESS_BITS: 64
CL_DEVICE_NAME: Quadro K610M
CL_DEVICE_VENDOR: NVIDIA Corporation
CL_DRIVER_VERSION: 353.90
CL_DEVICE_VERSION: OpenCL 1.2 CUDA
CL_DEVICE_MAX_COMPUTE_UNITS: 1
CL_DEVICE_LOCAL_MEM_SIZE: 49152
CL_DEVICE_MAX_WORK_GROUP_SIZE: 1024
CL_DEVICE_MAX_WORK_ITEM_SIZE: [1024,1024,64]
CL_DEVICE_IMAGE_SUPPORT: 1
====================================================

Global_work_items, width = 1920, height = 1120
Global_work_size cannot exceed CL_DEVICE_ADDRESS_BITS
       NDRange perf. counter time (100 Filter Calls) 2382.475260 ms.
Save Image: BinomialFilterOutput.bmp
```

When I test it on the target, there is no difference, I guess it may be the 2) queue lost some of waiting command because of some undefined bugs…on the laptop platform.