Realizing Multilayered Soft Tissue Model Using Multithreading in Python

Jayasudha.K1*, Mohan.G.Kabadi2

1*Research Scholar, VTU-RRC, Jnana Sangama, Belagavi, India
2Presidency University, Itgalpur, Bangalore, India

*Corresponding Author E-mail: 1jayanataraja@gmail.com, 2mohankg@presidencyuniversity.in

Abstract

Interactive Virtual Reality based surgical simulators are gaining popularity where an inexperienced surgeon can practice the surgery irrespective of time and cost which is an effective way in medical training. In order to guarantee the real time performance, a multithreading technique is applied to the model in different languages. This results in increased utilization of the processor resources and high update rate is achieved with a realistic rendering of the proposed graphics model for the virtual reality simulation. This paper presents a comparative analysis of the model’s performance using multithreading process. This is based on addition of matrices of the 3D model simulation with large dimensions in C and python programming languages implementations.

1. Introduction

In Virtual environments the main challenging task is to run the application model in real time that means how much time step is needed to simulate the application dynamically within a short period of time. As per industry standards performance increases with respect to higher clock speed. In fact even in desktop computers the computation power is increased by utilizing all computational resources to run parallel. In this paper, the comparative analysis of the performance were analyzed in different languages for the simulation of multilayer soft tissue modeling for virtual surgery application with special focus on parallel computations.

2. Method

Human skin is the single largest organ which is about 16% of human adult weight is not just a single layer of white membrane present beneath the skin. But is composed of multiple layers [1], with each layer clearly defined with proper functionality. As per figure 1, human skin is made up of three layers namely epidermis, dermis and sub-cutis layers. Epidermis is the topmost layer which is thin and smooth, acts as primary protective structure. Dermis is the second layer quite thick, also called as hypodermis. This contains only flesh and provides cushion effect. It is subdivided into two more sub layers called superficial dermis and deep dermis layers. Sub-cutis is the last layer which contains fat, connected to muscles and also supplies nutrients to top two layers. It is also called as hypo-dermis. Hence the core work is to simulate the prototype of multiple layers soft tissues that mimics the human skin. The prototype is developed by creating a cube data structure using three dimensional Delaunay triangulation techniques [2].The cube data structure consists of vertex-list, edge-list and triangle-list. The algorithm and data structure [3] satisfies the following criteria:

1 cube = 8 vertices + 18 edges + 12 triangles.
These cubes are placed with respect to 3D points along x-y-z axis on a plane, hence a 3D graphic image is realized on a 2D computer screen. To perform parallel computation these 3D points are analyzed as 2D matrices to perform basic addition operations on various programming platforms like C programming and python programming. In the next section a comparative study is performed to analyze the performance of matrix addition using C language implementation under sequential computing paradigm and python implementation under parallel computing paradigm.

3. C language Implementation

C is a basic procedural language that supports dynamic memory management providing several functions for memory allocation and management, these are found in stdlib.h header file. Their function and description are as listed below:

a) void *calloc (int num, int size): Allocates array of num elements each of which size is in bytes.

b) void free (void *address): Releases block of memory as specified by address.

c) void *malloc (int num): Allocates array of num bytes and leave them uninitialized.

d) void *realloc (void *address, int newsize): Reallocates memory to the specified address extending up to new size.

C library provides clock() function through header file <time.h> that returns number of clock ticks elapsed since the start of the program. To get number of seconds used by CPU it is required to divide by CLOCKS_PER_SEC. This is because on a 32 bit system CLOCKS_PER_SEC equals 1000000 and will return same value if divided by CLOCKS_PER_SEC. It is popular because of the following features:

- easy to learn- read- maintain
- broad standard library
- interactive mode
- portable
- extendable
- scalable
- databases
- GUI programming.

Python supports multi threaded programming that means running many threads concurrently. A thread is a light weighted process and do not require much memory. Multiple threads with in a process share the same data space. The thread class consists of several methods to implement threading, they are as follows:

- Run() : Entry point for a thread
- Start() : starts a thread by calling run()
- Join() : waits for threads to terminate
- isAlive() : checks whether thread is still executing
- getName() : return name of a thread
- setName() : sets name of a thread

One can define function in python to provide required functionality, it begins with keyword def followed by function name and parenthesis () and starts with a colon : and ends with return statement with an expression which is optional. The syntax is as follows:

def function-name(parameters):
    .......
    return [expression]

Python includes time module through header file <time.h> which provides functions for working with times. The function time.time () return the current system time. The algorithm for multi threading is as shown below.
Multithreading Algorithm:

1. Incorporate the threading module an API for spawning multiple threads
   ```python
   import threading
   ```
2. Define function with array of cubes
   ```python
def firstcube (nos):
    print ("The firstcube of given array")
    for n in nos:
        time.sleep (0.2)
        print ("firstcube n", arr1)

def secondcube (nos):
    print ("The secondcube of given array")
    for n in nos:
        time.sleep (0.2)
        print ("secondcube n", arr2)
   ```
3. Initialize 2D array elements of cubes
   ```python
   arr1=[[0,0,0],[20,0,0],[20,10,0],[0,0,20],[20,0,20],[0,10,20]]
   arr2=[[0,0,0],[40,0,0],[40,10,0],[0,0,20],[40,0,20],[40,10,20]]
   ```
4. Create threads
   ```python
t1 = threading.Thread (target=firstcube, args= (arr1,))
t2 = threading.Thread (target=secondcube, args= (arr2,))
   ```
5. Starting thread1 and thread2
   ```python
t1.start ()
t2.start ()
   ```
6. Wait for threads until completion
   ```python
t1.join ()
t2.join ()
   ```
7. Print threads after completion of execution with CPU time
   ```python
print (arr1 [:])
print (arr2 [:])
   ```

5. Results

To increase the performance of the model, the three layers of soft tissues with cubical data structure has been analyzed for parallel and serial computation in the form of addition of matrices with different dimensions. The simulation result tools used are as follows: dynamic programming using C compiler and multithreading using python compiler. Accordingly they are implemented in C and Python programming languages with computing environment as shown below:

**C Language computing environment:**

Intel®core(TM) i3 CPU M380@2.53Ghz
Ubuntu 16.04 64 bit OS, 2GB RAM
gcc version 4.8.4 – C compiler

**Python Language computing environment:**

Intel®core(TM) i3 CPU M380@2.53Ghz
VTK 6.1.0 Python 64-bit(Intel) on windows7
Python compiler

Matrix addition is performed by considering 3D points of all cubes along XYZ axis as a two dimensional matrix. In C programming matrix addition is performed directly with respect to number of cubes in an horizontal manner. In python matrix addition is performed treating each matrix as a single thread of execution in horizontal layered order. Table1 shows the comparisons of actual time taken for execution of addition of matrices in c and python language implementations. Dynamic programming in C allocates blocks of memory of any user defined size requested at run time using library function called `calloc ()`. These blocks can be reused or freed subsequently. Multithreading in python performs multi tasking i.e. parallelization of threads using functions: `start ()` and `join ()`. That means when one thread goes to sleep, the other thread will be active and vice versa towards completion of parallel execution. Figure 3 shows the graphical representation of program execution time. It is clearly understandable from table 1 that the execution time for python is almost same as in c for up to 80 cubes and then decreases gradually. This is because c utilizes heap for dynamic memory allocation, and stack for multithreaded situation such as python. Heap is application specific where as stack is thread specific. In heap memory allocation is at run time, where as in stack it is at compile time. In heap allocating and making free a block of memory is done at any time, so as the data increases it becomes more complex to keep track of the data and hence access to memory becomes a bit slower. And therefore takes more execution time.

<table>
<thead>
<tr>
<th>Sl No</th>
<th>No of Nodes</th>
<th>Time taken for execution(in sec) using C</th>
<th>Time taken for execution(in sec) using Python</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>0.846</td>
<td>1.888</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>1.223</td>
<td>1.956</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>1.737</td>
<td>2.168</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>2.577</td>
<td>2.464</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>4.564</td>
<td>3.339</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>4.724</td>
<td>4.275</td>
</tr>
<tr>
<td>7</td>
<td>80</td>
<td>5.574</td>
<td>5.023</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>8.948</td>
<td>6.24</td>
</tr>
<tr>
<td>9</td>
<td>120</td>
<td>11.036</td>
<td>7.036</td>
</tr>
<tr>
<td>10</td>
<td>140</td>
<td>11.576</td>
<td>7.986</td>
</tr>
<tr>
<td>11</td>
<td>160</td>
<td>13.32</td>
<td>8.553</td>
</tr>
<tr>
<td>12</td>
<td>180</td>
<td>13.524</td>
<td>10.217</td>
</tr>
<tr>
<td>13</td>
<td>200</td>
<td>15.737</td>
<td>11.513</td>
</tr>
</tbody>
</table>

Table 1: Comparison Table for Addition of Matrices
Where as in multithreaded situation, each thread will have independent stack and always reserved in LIFO (First In Last Out) order, so even though the data increases, keeping track and freeing a block is nothing compared to heap and hence access to memory is at a faster rate resulting in less execution time.

Table 2: Comparison Table for Binary Form Of Execution

<table>
<thead>
<tr>
<th>No of Nodes</th>
<th>using C</th>
<th>using Python</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.846</td>
<td>0.825</td>
</tr>
<tr>
<td>4</td>
<td>1.084</td>
<td>1.888</td>
</tr>
<tr>
<td>8</td>
<td>1.223</td>
<td>1.965</td>
</tr>
<tr>
<td>16</td>
<td>1.737</td>
<td>2.541</td>
</tr>
<tr>
<td>32</td>
<td>3.741</td>
<td>3.198</td>
</tr>
<tr>
<td>64</td>
<td>7.177</td>
<td>4.992</td>
</tr>
<tr>
<td>128</td>
<td>13.817</td>
<td>7.471</td>
</tr>
<tr>
<td>256</td>
<td>16.993</td>
<td>14.555</td>
</tr>
</tbody>
</table>

Table 2 shows the same comparison in C and python multithreading, but uses binary form of execution $2^n$ where $n$ is the number of cubes. Figure 4 shows the same in graphical representation. It is to be observed that for small number of cubes say about 32 the execution time is almost the same. And as the number of cubes increases, the execution time gradually decreases for python multithreading. Also to be noted that in multithreading environment the common recommendation is $n+1$ thread, where $n$ is number of CPU cores available. That means when $n$ threads work for CPU, one thread will be waiting for disk I/O. Hence having fewer threads will not utilize the CPU resource. Therefore more the number of threads will improve the performance by concurrently keeping the CPU busy.

This is also one of the reasons to be observed in graph that having less number of nodes say up to 80 in figure 3 and 32 in figure 4 the execution time is nearly the same as in c. So it can be analyzed and understood from the graph that the computation time taken for python multithreading implementation which works on parallel computation tremendously decreases with increase in number of cubes in comparison with C implementations which work on sequential computing environments.

6. Conclusion

In real time virtual environments it is essential that the performance of the models execution time is expected to be at a
higher rate for realistic rendering of graphics model such that proper utilization of resources takes place. The above mentioned implementation performs matrix addition operations under sequential and parallel computing paradigms. It can be concluded that many matrix operations on parallel computations like multithreading are fast and effective compared to sequential computing environments. To achieve much higher performance at a very fast rate and to solve many complex computational problems in an efficient way CUDA enabled GPU can be used for parallel computing applications as a future work.

References

[8] https://www.tutorialspoint.com/python/python_basic_syntax.htm