MotionTracking Grid Computing Guide

MotionTracking by
Yannis Kalaidzidis, Giovanni Marsico, Mikhail Chernykh
1997-2013

Manual by
Yury Bodrov and Mikhail Chernykh
Last Update: December 10, 2013

Contents

1 Grid Computing with MotionTracking

1.1 Introduction to Grid Computing

We live in an era of very large data sets, whether from high-throughput screens or movies taken at very high frame rates. In either case, researchers can easily generate thousands of images in a single day. Performing the Object Search routine on thousands of images is often more than one computer can handle. In order to handle very large data sets, MotionTracking is equipped to perform grid computing, whereby the program distributes the images to a networked cluster of computers. The same technique can be used to take advantage of dual-core and quad-core computers, by distributing separate image files to each of the processors for analysis. By distributing the computational tasks, the time required to process very large data sets is substantially reduced. The calculations can be done on remote PCs running windows, large super-computers/PC-Farms with Linux or other cores of the MotionTracking workstation itself. The generalized term for these in MotionTracking is "Application Server".

1.2 Configuring a Remote Machine for Grid Computing

In order for a computer to function in a computational cluster, it is necessary to install the Pluk development environment on the computer. Although it is not necessary to install MotionTracking iself, only the Pluk environment, Pluk is included in the current MotionTracking installation and it is recommended use it because of it's better installation wizard.

1.2.1 Installing Pluk

Download the latest MotionTracking installer from motiontracking.mpi-cbg.de and execute the downloaded setup file. Follow the instructions in the wizard. It is recommended to use the default C:/MotionTracking installation directory.

1.2.2 Creating a Pluk Account

First, make sure your MotionTracking installation directory contains a *dat* subdirectory, if not, create it. Execute *placclst.exe* that can be found in the *bin64* directory (if you only use 32 bit then use the *bin* directory). The account list is shared across both 32 and 64 bit versions so you don't need to repeat the procedure for 32 bit.
Create a new "Pluk account" under which you execute your remote calculations. The user name and password that you select here will be used later, when you configure Motion Tracking to connect to the remote computer. The Pluk Account List should appear as shown in Figure 1. After creating a new account, you must restart the Pluk Services Launcher, as described above, by either restarting the computer or by stopping and restarting the "Pluk Services Launcher" in "Services" of the Windows Task Manager.

This procedure will enable you to calculate on one single core of the machine, but most likely you will want to use multiple cores. If the initial username is "username" you need to simply add another user with the name "username1" for the second core, then "username2" for the third and so on. Note that they all have to have the same password. See Figure 1 for an example.
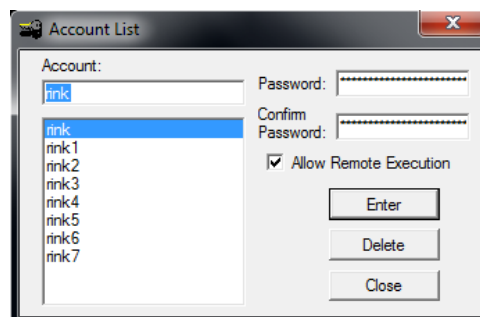
Figure 1: Creating a Pluk Account. In order to create an account, enter the user name (e.g., rink), enter and confirm a password, and select "Allow Remote Execution". Pressing the Enter button will cause the new account to appear in the Account List. In this example the machine is set up to execute calculations on 8 cores.

Warning: if the account list dissapears if you close and reopen *placclist.exe* you need to create a folder "dat" in your MotionTracking installation directory.
(e.g. C:/MotionTracking/dat)

1.2.3 Launching Pluk Services: plbrok and rdaemon

It is necessary to have both *plbrok.exe* and *rdaemon.exe* services running on the application server to calculate tasks. The most elegant way is to create a Windows batch file that will start both services for you. Just create a new file with the extension .bat in the bin or bin64 folder inside you MotionTracking installation directory, for example "start_rdaemon.bat" and put the following two lines in it:

```
start plbrok.exe /hide
start rdaemon.exe /hide
```

Now double click on that file to start both sevices at once. If you want the machine to be available for remote calculations at all times, put a shortcut to this file in the startup directory of the start menu.

1.3   Configuring MotionTracking for Grid Computing

Now that the remote computer is set up properly, we are ready to configure Motion Tracking on your machine to connect to the remote computer and distribute computational tasks.

First of all, Start Motion Tracking.

1.3.1   Adding a remote computer to the list of available servers

Go to "Options → Grid Computing". This command brings up a dialog box that lists the servers available for grid computing. This dialog box appears in Figure 2. Right-click in the white space of the Server List, select Add in order to add a new server.
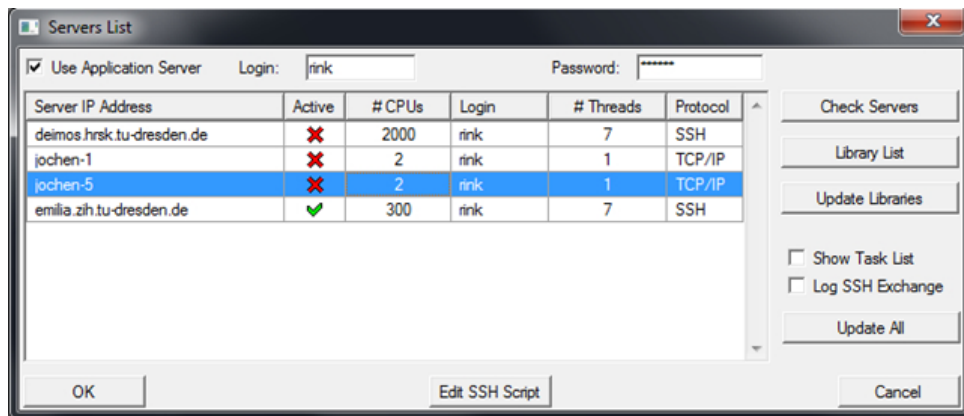

Figure 2: Server List for Motion Tracking. The grid computing option is turned on by checking the "Use Application Server" box. In this example we have two PC farms, deimos and emilia, connected by SSH and two remote machines, jochen1 and 5, connected by TCP/IP.

Right-click in the white region of the Server list, select Add. This will create a new line in the list of servers, where you can add the IP address of the new computer. The IP address can be either the name of the remote computer on the local area network (LAN) (e.g., remote-computer-12) or the IP address of the computer (e.g., 10.24.217.121).
The "#CPUs" is the number of tasks you want to send to the remote machine for simultaneous calculation. It is advised **not** to enter a number greater than the number of cores you have set up on the remote machine because the queuing over TCP/IP is not that good. You can use any number if you use SSH on PC Farms because they have their own queueing systems. Select TCP/IP in the last column for remote machines and SSH for PC Farms (more on those later).

The Login and Password should be filled in with the same login and password given to the remote computer, above. You may use a different login and password for each server added to the list.

1.3.2   Verifying the Library List

Click on Library List. This will open the dialog box. The following three libraries should be present: math.dll, cmath.dll, and imageframe.dll. Make sure that all three libraries are checked with a green check-mark. Click Update Libraries to confirm.

1.3.3   Activating Application Server

When the "Use Application Server" button is selected, Motion tracking will distribute the computational tasks to the servers in the Server List. Each server in the list has an Active or Inactive status, shown by a green check-mark or a red X. By toggling the status of the different servers in the Server List, you can distribute the computation to only a subset of the available servers.

1.3.4 Checking Servers

After this is done you can click on "Check Servers". If one of the active servers is not working it will deactivate it. Double-Check your data in this case, make sure the remote computer is running and check if the firewall settings allow the communication.

1.3.5 Configuration of a multi-core Motion Tracking Workstation

Most computers nowadays have multiple cores. The MotionTracking program itself uses only one core, but it can send computational tasks to all cores of the workstation.

If the computer running Motion Tracking (not a remote computer, but the computer with the full program) has a multi-core processor, it can be set up in a similar way to remote machines. First of all you need to create a pluk account on your own machine as described previously in section 1.2.2. Then you need to open the Server List by going to "Options → Grid Computing". Add a new server with the IP address: 127.0.0.1. This IP address is used by every computer to address itself. Change the #CPUs column to the number of cores on your machine that you want to use simultaneously. No login or password are necessary for configuration of a computer with IP address: 127.0.0.1. Selecting Use Application Server will now cause Motion Tracking to send one image to each core of the multi-core computer for processing.

1.4 Connection to a PC-Farm over SSH and psftp

First of all you must consult the administration of the PC-Farm for login information and the authorization files (id_rsa.ppk from step 7) and help with SSH if necessary.

It is also assumed that MotionTracking is installed on PC-Farm since this is a complicated process. As of now, MT is only installed on the PC-Farms Atlas and Emilia of the "center for information services and high performance computing" (ZIH) in Dresden.

You will need to install several third party programs to get the system up and running on your local workstation. When you see strings like "...\OpenSSH\bin" it means that you have to type the full path to you OpenSSH install directory instead of the "...".

1. Download Cygwin from cygwin.com

2. Install it to "C:\Program Files\OpenSSH". This is non-optional. If you use a non-English windows and Program Files doesn't exist, create it.

3. The installation is a bit tricky. Proceed until you get to the window shown above. Meanwhile you will be asked where you want to download the packages and other stuff, most of which is self explanatory. Select http on inf.tu-dresden.de as download server if in doubt.
When you reach the window where you have to select the packages that will be installed search for ssh and select the OpenSSH package by checking the right checkbox in the Bin column. You don't need to select the Source Checkbox. Hit next when done and proceed. It will ask you if you want to install automatically selected additional packages that OpenSSH needs to run which you must do. Finish the installation.

4. Download psftp and putty (http://www.chiark.greenend.org.uk/ sgtatham/putty/). (Encryption is legal in Germany btw.)

5. Copy psftp.exe to "...\OpenSSH\bin" folder. Make sure that you install the Pagent and Puttygen components as well.

6. Run ssh session from terminal. To do that open the terminal (run cmd) and type "ssh user@example.server.com" with the correct information for the pc farm that you want to connect to.

7. Exit ssh session by either closing the terminal or typing "exit" or "logout" (depends on the machine you are logged on)

8. Optional: Download and install WinSCP (http://winscp.net/eng/download.php) for an easy way to do get the next few steps done. If you think that you can do that with psftp and the command line, you can do that of course but it is way more convenient to do it with WinSCP.

9. Copy "id_rsa.ppk" from your pc-farm node to "...\OpenSSH\home\ [your windows username] .ssh" and to "...\OpenSSH\bin". To do that use WinSCP and login onto your PC-Farm with it.

10. Test connection (Open Terminal (run cmd),
C:\...\OpenSSH\bin\ssh farm_user@farm.example.com). If it asks for a password, something is wrong. Most likely it is the permissions of the files that were copied into the .ssh directory, because windows resets them while copying. Therefore you need to use the cygwin terminal that was installed in one of the earlier steps and type:

```
cd .ssh
chmod 600 *
cd ..
chmod 700 .ssh
```

11. Test psftp connection by using the following command from the terminal:
"...\OpenSSH\bin\psft.exe -i ...\OpenSSH\bin\id_rsa.ppk". This should go through without any password requests as well.

12. Note that on a x64 system you must *copy* the Cygwin installation from "C:\Program Files\OpenSSH" to "C:\Program Files (x86)\OpenSSH", otherwise the 32bit version will not work.

Once done, add the Farm to the server list the same way you would do with a remote computer.

Warning: SSH must go through without any password requests. Use proper SSH authentification files. psftp should connect without any further requests as well if called with an additional flag as seen in the example below. You can check both by trying both out from the windows command line:

```
ssh loginname@server.example.com
psftp loginname@server.example.com -i ..\OpenSSH\bin\id_rsa.ppk
```

1.5   Using the Application Server

After the remote computer and the Motion Tracking workstation have been configured for Grid Computing, and you have selected the Use Application Server option, simply start any calculation procedure which supports Grid Computing you normally would.

For example, select "Object → Object → Find Objects" and select the "Use Gird Computing". Starting the Object Search procedure with the Application Server active will bring up a dialog window shown in Fig. 3:

Click the Start button to begin the computation. The window will specify the specific server to which each image has been sent, the total processing time, the number of objects found in each image, etc. You can stop the computation by pressing the Stop button.

If you want to change servers during running calculation, hit the stop button, click on "list servers", select the servers you like or change any options. When done, simply start again. Any dispatched tasks will be resent if the server that they were sent on was either deactivated or its options changed.

If you encounter any errors due to any reason you have the option to resent all failed tasks by clicking the "Rebuild Failed" button. The "List Failed" button will give you a popup window with a list of all tasks which have the status "failed". "Reset Dispatched Tasks" does what the name suggests - a task counts as dispatched when it was sent to the remote machine (Server Address and Dispatch Time columns are filled) but it's state is neither "completed" nor "failed".

Each of these functions is accessible for individual task by right-clicking on the task.

MotionTracking is designed to handle huge amounts of data. Since this is far too much to handle most of the personal computers, all of the computation-heavy calculations can be out-sourced to grid computers and pc-farms.

- object search

Figure 3: Grid Computing window

- per frame correction

- track search in batch mode

- statistic calculation in batch mode

1.6 Troubleshooting

1.6.1 Checking the state of the Pluk Service Launcher

Go to "Start → Control Panel → Administrative Tools → Services". This will open the Services list for your Windows computer. The list of Services will include Pluk Services Launcher. The start-up type for this service should be set to *Automatic*. If you restart the machine after installation this should be the case anyway.

1.6.2 Verifying that the Pluk Services are active

Go to "Start → Programs → Pluk → Pluk Services List". This will bring up a list of the services that Pluk uses in the operating system. There are two services required for grid computing: plbrok and rdaemon. Check that both of these services are visible in the Pluk Services List. If one of the services is missing, go to file menu: "Service→Add". The "Add Service" command will create a new line in the Services list, into which you can enter the information shown in 4. After adding any missing services, go to the file menu: "File→Save" to save the newly added Services.

Select Auto-start for both services. After adding the two services, it is necessary to activate them on your machine. This can be done by (a) restarting the computer, or (b) restarting the Pluk Services Launcher. To restart the Pluk Services Launcher, go to "Start→Control Panel→Administrative Tools→Services". Select the Pluk Services Launcher, and stop the service (right-click to bring up a context menu, or press the Stop button on the top menu bar), then restart the service (again, right-click or use the Start button to the top menu bar).

Once the Pluk Services Launcher has been restarted, you can verify that the two processes are operating in memory by looking at the Processes list for your machine. To bring up the list of active processes, press Ctrl-Alt-Delete. This opens the Windows Task Manager. The Task Manager has a tab called Processes, which lists all of the active processes currently running in memory. In order for grid computing to function on a computer in a computational cluster, the following services must be running: plbrok.exe, plukserv.exe, rdaemon.exe, and rserver.exe. Please verify that all four services appear in the Windows Task Manager before proceeding.
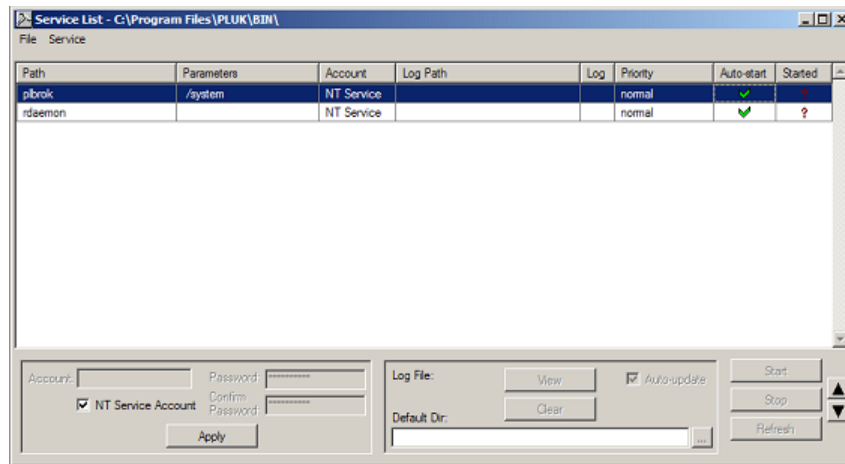
Figure 4: Pluk Services List. The dialog box above shows the services as they should appear for Grid Computing. If either of the two services are missing, they can be added as described above.

Important Note: if the remote computer is separated from your Motion Tracking workstation by a firewall, it is required that the Pluk processes are able to communicate through the firewall. Contact your network administrator and grant firewall access to the processes: plbrok.exe and rserver.exe. Without firewall access, the grid computing features will not function.