QLWFPC2: Parallel-Processing Quick-Look WFPC2 Stellar Photometry based on the Message Passing Interface

Kenneth John Mighell
National Optical Astronomy Observatory
mighell@noao.edu

ABSTRACT

I describe a new parallel-processing stellar photometry code called QLWFPC2 which is designed to do quick-look analysis of two entire WFPC2 observations from the Hubble Space Telescope in under 5 seconds using a fast Beowulf cluster with a Gigabit Ethernet local network. QLWFPC2 running on 4 processors takes about 2.4 seconds to analyze HST WFPC2 archive observations of M54 (NGC 6715) which is the bright massive globular cluster near the center of the nearby Sagittarius dwarf spheroidal galaxy. The analysis of these HST observations of M54 lead to the serendipitous discovery of more than 50 new bright variable stars in the central region of M54 where no variables have been reported by previous ground-based studies of variables in M54. This discovery is an example of how QLWFPC2 can be used to quickly explore the time domain of observations in the HST Data Archive. Further information about QLWFPC2, including documentation and source code may be found at the following web site: http://www.noao.edu/staff/mighell/qlwfpc2

MOTIVATION

Software tools which provide quick-look data analysis with moderate accuracy (say 3 to 6 percent relative precision) could prove to be very powerful data mining tools for researchers using the U.S. National Virtual Observatory (NVO). The NVO data server may also find quick-look analysis tools to be very useful from a practical operational perspective. While quick-look stellar photometry codes are excellent tools to create metadata about the contents of CCD image data in the NVO archive, they also can provide the user with real-time analysis of NVO archival data. It is significantly faster to transmit to the NVO user a quick-look color-magnitude diagram (consisting of a few kilobytes of graphical data) than it is to transmit the entire observational data set which may consist of 10, 100, or more megabytes of data. By judiciously expending a few CPU seconds at the NVO data server, an astronomer using the NVO might well be able to determine whether a given set of observations is likely to meet their scientific needs. Quick-look analysis tools thus could not only provide a better user experience for NVO researchers but it could simultaneously allow the NVO data servers to perform their role more efficiently with better allocation of scarce computational resources.

Successful quick-look analysis tools must be fast. Such tools must provide useful information in just a few seconds in order to be capable of improving the user experience with the NVO archive.

QDPHOT

The MXTOOLS package for IRAF (http://www.noao.edu/staff/mighell/mxtools) has a fast stellar photometry task called QDPHOT (Quick & Dirty PHOTometry) which quickly produces good (~5 percent relative precision) CCD stellar photometry from 2 CCD images of a star field. For example, QDPHOT takes a few seconds to analyze 2 Hubble Space Telescope WFPC2 frames containing thousands of stars in Local Group star clusters (Mighell 2000: ADASS IX). Instrumental magnitudes produced by QDPHOT are converted to standard colors using the MXTOOLS task WFP2COLOR.

QLWFPC2

I have recently implemented a parallel-processing version of the combination of the QDPHOT and WFP2COLOR tasks using the MPICH implementation of the Message Passing Interface (MPI) from the Argonne National Laboratory (http://www-unix.mcs.anl.gov/mpip1/manip) . This new stand-alone multi-processing WFPC2 stellar photometry task is called QLWFPC2 (Quick Look WFPC2) and is designed to analyze two complete WFPC2 observations of Local Group star clusters in under 5 seconds on a 5-node Beowulf cluster of Linux-based PCs with a local network based on a Gigabit Ethernet router. QLWFPC2 is written in ANSI C and uses the CFITSIO library (http://heasarc.gsfc.nasa.gov/docs/software/fitstools), from NASA’s Goddard Space Flight Center, to read FITS images and the Parameter Interface Library (http://isdc.unige.ch/bin/std.cgi?Soft/isdc_releases_public/osa-2.9), from the INTEGRAL Science Data Center, for the IRAF parameter-file user interface.

QLWFPC2 PERFORMANCE

The current implementation of QLWFPC2 was tested on a Beowulf cluster composed of 5 single 1.8-GHz AMD Athalon CPUs with 3 GB total memory connected with a Gigabit Ethernet router with 40 GB of local disk and 120 GB of NFS-mounted disk. QLWFPC2 running on 4 processors takes ~2.4 seconds (see Figure 1) to analyze WFPC2 archive data sets u37ga407r.c0.fits (F555W; 300 s) and u37ga407r.c1.fits (F814W; 300 s) of M54 which is the bright massive globular cluster near the center of the Sagittarius dwarf spheroidal galaxy. QLWFPC2 analyzed over 50,000 point source candidates and reported 1, I, F555W and F814W photometry of 14,611 stars with signal-to-noise ratios of 8 or better. The analysis of these HST observations of M54 lead to the serendipitous discovery of more than 50 new bright variable stars in the central region of M54 where no variables have been reported by previous ground-based studies of variables in M54.

RECOMMENDATIONS:

- Buy fast machines. QLWFPC2 almost met the design goal of 5 seconds with a single CPU. Note that an infinite number of machines operating at ~1 GHz could not meet the 5-s design goal.
- Buy fast networks. Gigabit Ethernet is ideally suited for today’s GHz-class CPUs and is now very affordable. Old networks operating at Fast Ethernet speeds will be bandwidth-bound for tasks requiring large (>1 MB) messages. The test Beowulf cluster has a latency of 90 microseconds and a sustained bandwidth of 33 MB/s for large messages.
- Buy fast disks. The main disk of the test Beowulf cluster can read large FITS files at a respectable ~30 MB/s with 7200 rpm disks. Nevertheless, reading 2 WFPC2 images still takes 0.6 s to read which is a significant fraction of the measured total execution times.

ACKNOWLEDGEMENT

This work is supported by a grant from the National Aeronautics and Space Administration (NASA), Interagency Order No. 5-13811-G, which was awarded by the Applied Information Systems Research Program (AISR) of NASA’s Office of Space Science (NRA 01-098-01).