NSCEE
National Supercomputing Center for Energy and the Environment
High-Performance Computing and Communications in Nevada

January, 2004
Vol. 14, No. 1

TeraWord

The Newsletter of High-Performance Computing and Communications in Nevada

Contents

Cluster Donation: HP Identifies NSCEE as Itanium 2 Reference Site

Hewlett-Packard (HP) announced in December 2003 the donation of a rx2600 Integrity Cluster to the University of Nevada Las Vegas, National Supercomputing Center for Energy and the Environment.

HP awarded a similar donation to four other universities including: UC San Francisco, USC, UC San Diego and Oregon State. The primary purpose of the donation will be to evaluate performance and functionality for the migration of legacy applications.

HP is hopeful that this contribution to NSCEE will facilitate NSCEE's responsibilities to provide supercomputing training and services, maintain and plan the overall information technology architecture, support user tools, and develop new software technologies. More importantly, the new technology will be a significant enabler of immerging scientific research for NSCEE's user community.

The Intel Itanium 2 processor is a milestone in the continuing evolution of microprocessors because it is the first enterprise-class 64-bit processor that has the power to become pervasive (see related article on page 3).

Today, 32-bit servers (i.e., Intel's Pentium) or proprietary 64-bit RISC servers are the norm. They have respectable price/performance ratios but are either fundamentally limited in performance scalability or are exceedingly expensive.

Servers based on the Intel-architecture 32-bit (IA-32) processors (i.e., Pentium), for example, are unable to address large amounts of memory efficiently. Meanwhile, 64-bit RISC architectures have the necessary performance and addressing, but they are more expensive and may lock the user into a proprietary operating environment and a single computer vendor. End users and IT professionals alike are clamoring for high performance and large addressing at an economical price. Additionally, they are asking for a pervasive environment that avoids the complexity of dealing with multiple architectures.

continued on page 4

rx2600

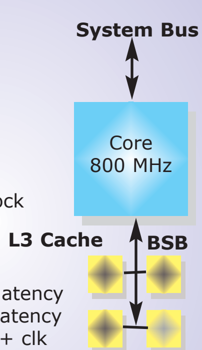
Itanium Processor

System Bus
64 bits wide
133MHz/266 MT/s*
2.1 GB/s

Width
2 bundles per clock
4 integer units
2 load or stores per clock
9 issue ports

Caches
L1 - 2x16KB - 2 clock latency
L2 - 96K - 6.9+ clock latency
L3 - 4MB external - 20+ clk
11.7GB/s bandwidth

Addressing
44 bit physical addressing
50 bit virtual addressing
Maximum page size of 256MB



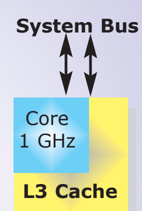
Itanium 2 Processor

System Bus
128 bits wide
200MHz/400 MT/s*
6.4 GB/s

Width
2 bundles per clock
6 integer units
2 load and stores per clock
11 issue ports

Caches
L1 - 2x16KB - 1 clock latency
L2 - 256K - 5+ clock latency
L3 - 1.5MB - 3MB - 12+ clk
32GB/s bandwidth

Addressing
50 bit physical addressing
64 bit virtual addressing
Maximum page size of 4GB



Itanium 2 Enhancements

- **3x improvement in FSB bandwidth**
- **2x improvement in cache latencies and bandwidth**
- **More units for execution of integer code**

* MT/s = Mega-Transfers/second

Research Activities at NSCEE

Ongoing Research on DOE Medical Records

The problem of handwriting recognition had surfaced even before the advent of computers. In 1929 Tausheck obtained a patent for an Optical Character Recognition (OCR) device using optical and mechanical template matching. After the dawn of the computer era the field progressed by leaps and bounds and, although the problem of recognizing human handwriting remains largely unsolved, OCR flourished and expanded into a multi-disciplinary science spanning such diverse fields as image processing, pattern recognition, differential topology, and stochastic processes to name just a few.

A DOE medical record form may not be an object of fascination to many, but to the Digital Image Processing team at UNLV, headed by Professor Angelo Yfantis, these records are fertile incubators of many interesting challenges in handwriting recognition and document analysis. Researchers Jerry Derby, Dimitri Papaioannou, John Istle, and Meenakshisundaram Murugan of the Image Processing Lab have dived into the problem of DOE medical records processing, combining an arsenal of traditional and novel techniques. The strategies employed by the team broadly fall into four categories: pattern matching, structural analysis, adaptive learning, and contextual verification.

In pattern matching techniques the object of interest is matched against different known patterns in an attempt to identify the closest fit. Since hand-written characters do not obey rigid structural principles, significant intelligence must be introduced in the matching process. For example, variations in rotation and size must be accounted for and eliminated, as well as superfluous strokes.

Structural analysis covers a wide range of techniques that attempt to decompose single handwritten characters in basic structural characteristics, called primitives. Primitives can be topological such as the number of holes, geometrical such as angles and curves, analytical such as length and density, graph-theoretical such as end-points and joins.

Adaptive learning is a rich and vast array of techniques used in pattern classification problems that include Neural Networks, Fuzzy Logic, Linear Regression, Genetic Algorithms, and more. The general idea behind these methods is that the system is "learning" to recognize a given pattern by examining several samples of this pattern. This process is adaptive in the sense that as the system obtains more samples, it becomes a more efficient system. The general success of these methods notwithstanding, they are not a recognition panacea. Deep insight into the problem at hand is required in identifying appropriate characteristics that will be beneficial to learning. Digitized images of handwritten characters are rich in variation, as can be seen, for example, in figure1. It is necessary therefore to extract quantitative characteristics that describe the essence of the character and do not depend on the writing style. This process is called feature extraction and it is similar to the use of primitives in structural analysis.

Contextual verification uses broader sources of information to improve the results of recognition. For example, in English word recognition, knowledge of the grammar structure and letter frequencies can be used for sanity checking. In Social Security Number recognition, access to an employee database can identify and fix an incorrectly identified number.

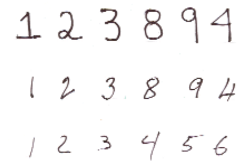


Figure 1. Samples of handwritten numerals

Researchers at the Image Processing Lab refine and combine the above approaches in hybrid methodologies to improve the accuracy of the recognition task.

Some of the accomplishments of the team to date include:

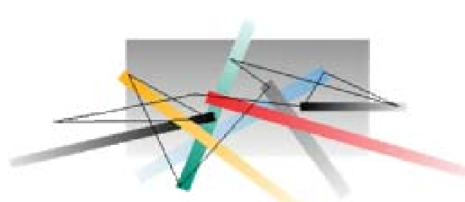
- identification of interesting portions of the form, such as the Social Security Number box
- recognition of hand-written social security numbers with database backup for contextual verification
- identification of checkboxes and whether they are checked or not checked
- form identification by locating and recognizing the logo
- separation of handwritten text from typewritten text including removal of lines
- document processing algorithms including segmentation of the form in sections and lines, and noise removal
- segmentation of loosely connected characters using character width and space width statistics.

Though the important problems of form and patient identification are addressed, many interesting challenges lie ahead. To name a few, the translation of the patients medical record from paper to electronic form, which involves decoding a combination of printed and hand-written text, compensation for common spelling errors, prescription identification - involving the Herculean task of deciphering doctors' handwriting - and signature verification. Armed with experience and a rapidly growing OCR toolkit, the Digital Image Processing team is up to the challenge.

For additional information, contact Dimitri Papaioannou at dimitri@cs.unlv.edu.

NSCEE Celebrates 14th Year as Member of CASC

The Coalition for Academic Scientific Computation (CASC) is a nonprofit organization of supercomputing centers and research universities that offer leading edge hardware, software, and expertise in high performance computing resources and “advanced visualization environments.”



Founded in 1989, CASC has grown into a national association representing 36 centers and programs in 22 states. NSCEE has been a member of CASC since 1990.

Working individually and together, coalition members complement traditional methods of laboratory and theoretical investigation by using high performance computers to simulate natural phenomena and environmental threats, handle and analyze data and create images - all at performance levels not available from smaller computers.

By applying advanced technology, CASC members help extend the state of the art to achieve the scientific, technical, and information management breakthroughs that will keep the U.S. in the forefront of

continued on page 4

Help Desk

Clear Advantages to 64-bit Computing

64-bit processors DO make a difference for two reasons: memory and processing power. Today's 32-bit chips work with 32 bits of data at a time and can address up to 4GB (approximately 2^{32} bits) of memory. In Windows-based machines, that 4GB is split between the operating system and the applications. That means the most memory any given application can access is 2GB. Intel's new Itanium 2 64-bit processor, by contrast, works with 64 bits of data at a time and can address up to 16 terabytes (approximately 2^{64} bits) of memory. The new processors should dramatically increase processing speed for complex math and graphics applications.

The blatant advantage of a 64-bit processor is high precision mathematics. The integer range that a 32-bit processor can handle is -2.1 billion to 2.1 billion. In other words, it can handle a number with nine significant figures. Tricks can be used to deal with larger numbers that amount to multiple memory addresses for each value and use advanced functions available in the programming compilers. **Tricks are useful, but not fast.**

Advanced math applications are found in large financial systems, computer simulations, CAD/CAM workstations, graphics rendering, and more importantly, encryption. In the networked world we live in, encryption is commonplace and rapidly expanding. Half of today's digital security system is based on the algorithm used to encrypt the data; the other is the size of the keys used to archive and extract the data. A strong algorithm with a weak key can be defeated by raw brute force in a short time, so large keys are a necessity. Today, with 32-bit processors, a strong key is 256-bit requiring eight addresses per value (8 x 32 bit = 256 bit) and lots of mathematical tricks. A 64-bit processor will use only four addresses per value (4 x 64 bit = 256 bit) for the same size key and significantly increase the speed of the encryption processes.

Extra speed will let programmers add remarkable detail, better textures, more realistic sounds, and larger and more realistic environments to their applications. Plus, the characters themselves will be rendered with dramatically more detail. You'll see more realistic representation of features such as hair, skin, and eyes.

The processor isn't the only player here, the operating system and applications must be tweaked (at a minimum, recompiled) to take advantage of the additional power.

The 64-bit processors from Intel (Itanium 2), AMD (Athlon64) and Apple (G5), will be significantly more expensive than those available in the past from vendors such as DEC and Sun. And applications for these new processors may be quite limited for some time to come, or limited to the same applications already available on UNIX machines running on other 64-bit processors.

If you are primarily a Microsoft Office user, you probably aren't bumping up against any of the limits of 32-bit processors. But if you're running scientific or graphics applications on a workstation, the improved graphics capabilities of 64-bit should interest you.

NSCEE Hosts 2nd Annual IT Workshop for the National Defense University (NDU)

The NDU Advanced Management Program (AMP) is a 14-week, in-residence, graduate-level program for senior military, federal civilian, and industry leaders focused on preparing them for the role of Chief Information Officer (CIO) within their organization. The capstone event is a week-long field study in a chosen domestic location, to engage with CIOs in government, industry, and academic settings, and learn first-hand about their information technology management challenges and best practices. Upon return to Washington, DC, the students must present their analysis and synthesis of the experience to the entire faculty of the Information Resources Management College (IRMC), one of the three colleges operating under the aegis of the National Defense University.

NDU identified NSCEE as one of its field study sites due to its unique IT environment and the support it provides to government, industry and academia. Joseph Lombardo, Director NSCEE, served as guest speaker. In addition to NSCEE, speakers (at their respective sites) included CIOs from Hoover Dam, Harrahs, Nellis Air Force Base, University Medical Center- Trauma and Children's Center, UNLV Office of Information Technology, the City of Las Vegas, and the State of Nevada.

The NSCEE workshop was held on Friday, December 5th for 19 NDU students and faculty representing the following organizations: Department of the Army and Army Reserves, Environmental Protection Agency, Raytheon Corporation, Philippine Air Force, Defense Intelligence Agency, National Reconnaissance Office, Office of the Joint Chiefs of Staff, State Department,

Visiting NDU students and faculty with Access Grid (see July 2003 issue of TeraWord) in background

Defense Contract Management Agency, Israeli Army and Bulgarian Army.

NDU, founded in 1976, is the premier center for joint professional military education and is under the direction of the Chairman of the Joint Chiefs of Staff. The University is located at Ft. McNair in Washington, DC. NDU's Joint Forces Staff College is in Norfolk, Va. NDU is accredited by the Commission on Higher Education of the Middle States Association of Colleges and Schools. Additional information may be found at www.ndu.edu.

Cluster Donation
continued from page 1

HP and Intel co-developed the Intel Itanium 2 microarchitecture. The result is a high-performance, parallel 64-bit architecture that has the performance headroom to grow in the future. The Intel Itanium 2 processor is the fundamental building block of NSCEE's HP Integrity server, the rx2600.

NSCEE Celebrates 14th Year as Member of CASC
continued from page 2

the 21st century information technology revolution.

Coalition members are involved in activities that foster major advances for virtually every element of society. The range of these efforts encompasses:

- Aiding in Homeland Security
- Accessing Information
- Improving Health Care
- Conducting Research
- Combatting Cyber-Terrorism
- Enhancing Education
- Supporting the Arts
- Innovating in Design and Construction
- Understanding the Environment
- Preparing for Bio-Terrorism
- Advanced Bioinformatics

More information about CASC and its membership may be found at www.casc.org.

NSCEE History - an image from our past

May 1990 - The Cray DD-49 Disk Pack (Disk Drives)

The Cray DD-49 Disk Pack was comprised of 5 individual drive units (4 operational plus one spare), each with a capacity of 1250 Megabytes (1.2 GB). The 5 units side by side took up about 10 feet in length and 4 feet in width ... all that floorspace for just 4.8 GB of disk storage! Today, a 181 GB drive can just about fit into the palm of your hand!

Visit http://www.sgi.com/pdfs/3504.pdf for SGI's writeup of the NSCEE Succcess Story, "Instant Multi-Platform Data Sharing Accelerates Projects."

Articles Invited
The National Supercomputing Center for Energy and the Environment invites you to contribute articles on your work on high-performance computers and especially our resources. Please submit your articles to:

TeraWord
UNLV/NSCEE
4505 Maryland Parkway
Box 454028
Las Vegas, NV 89154-4028

email
teraword@nscee.edu

Phone
(702) 895-4153

Fax
(702) 895-4156

NSCEE
National Supercomputing Center for Energy and the Environment
High-Performance Computing and Communications in Nevada

National Supercomputing Center for Energy and the Environment
4505 Maryland Parkway, Box 454028
Las Vegas, NV 89154-4028

Visit us at **www.nscee.edu**

UNLV
UNIVERSITY OF NEVADA LAS VEGAS