CESDIS

Center of Excellence in Space Data and Information Sciences

Annual Report

Year 11

July 1998 - June 1999

Dr. Yelena Yesha, Director

(submitted by Dr. Susan Hoban, Acting Director)

CESDIS
NASA Goddard Space Flight Center
Code 930.5
Greenbelt, MD 20771

phone: 301-286-4403
fax: 301-286-1777
e-mail: cas@cesdis.gsfc.nasa.gov
http://cesdis.gsfc.nasa.gov

Operated by Universities Space Research Association in cooperation with the National Aeronautics and Space Administration

FOREWARD

This report summarizes the range of computer science-related activities undertaken by CESDIS for NASA in the twelve months from July 1, 1998 through June 30, 1999. These activities address issues related to accessing, processing, and analyzing data from space observing systems through collaborative efforts with university, industry, and NASA space and Earth scientists.

The sections of this report which follow, detail the activities undertaken by the members of each of the CESDIS branches. This includes contributions from university faculty members and graduate students as well as CESDIS employees. Phone numbers and e-mail addresses appear in Appendix F (CESDIS Personnel and Associates) to facilitate interactions and new collaborations.

TABLE OF CONTENTS

OVERVIEW

CESDIS, the Center of Excellence in Space Data and Information Sciences, was developed jointly by the National Aeronautics and Space Administration (NASA), Universities Space Research Association (USRA), and the University of Maryland in 1988. It is operated by USRA, under a contract with NASA. The program office and a small, core staff are located on site at NASA's Goddard Space Flight Center in Greenbelt, Maryland.

USRA and the CESDIS Science Council

USRA is a nonprofit consortium of 80 colleges and universities, offering graduate programs in space sciences or related areas, which operates research centers and programs at several NASA centers. Most notable are the Lunar and Planetary Institute (LPI) at the Johnson Space Center in Houston, Texas, the Institute for Computer Applications in Science and Engineering (ICASE) at the Langley Research Center in Hampton, Virginia, the Research Institute for Advanced Computer Science (RIACS) at the Ames Research Center at Moffett Field, California, and the Stratospheric Observatory for Infrared Astronomy (SOFIA) in Waco, Texas.

Oversight of each USRA institute or program is provided by a science council which serves as a scientific board of directors. Science council members are appointed by the USRA Board of Trustees for three-year terms. Members of the CESDIS Science Council during 1997-1998 were:

- Dr. Rama Chellappa
  University of Maryland College Park
- Dr. Burt Edelson
  George Washington University
- Dr. Richard Muntz
  University of California, Los Angeles
- Dr. David Nicol
  Dartmouth College
- Dr. Jacob Schwartz
  New York University
- Dr. Harold Stone (Convener)
  NEC Research Institute
- Dr. Satish Tripathi
  University of California, Riverside
- Dr. Mark Weiser
  Xerox PARC

The CESDIS Science Council meets annually at Goddard to review ongoing CESDIS research programs and new initiatives.

The CESDIS Mission

The CESDIS mission is to increase the connection between computer science and engineering research programs at colleges and universities and NASA groups working with information science and technology applications in earth and space sciences. CESDIS also focuses attention on information science issues involved in storing, accessing, processing, and analyzing data from space observing systems, and collaborates with NASA space and earth scientists in research related to NASA's needs.

The CESDIS Seminar series seeks to offer the Goddard information science and technology community an opportunity to consider and discuss interesting advances in information science.

To suggest a speaker for the Spring 2000 series, please contact Susan Hoban.

CESDIS World Wide Web Homepage

The CESDIS web site is fully indexed and can be located through:

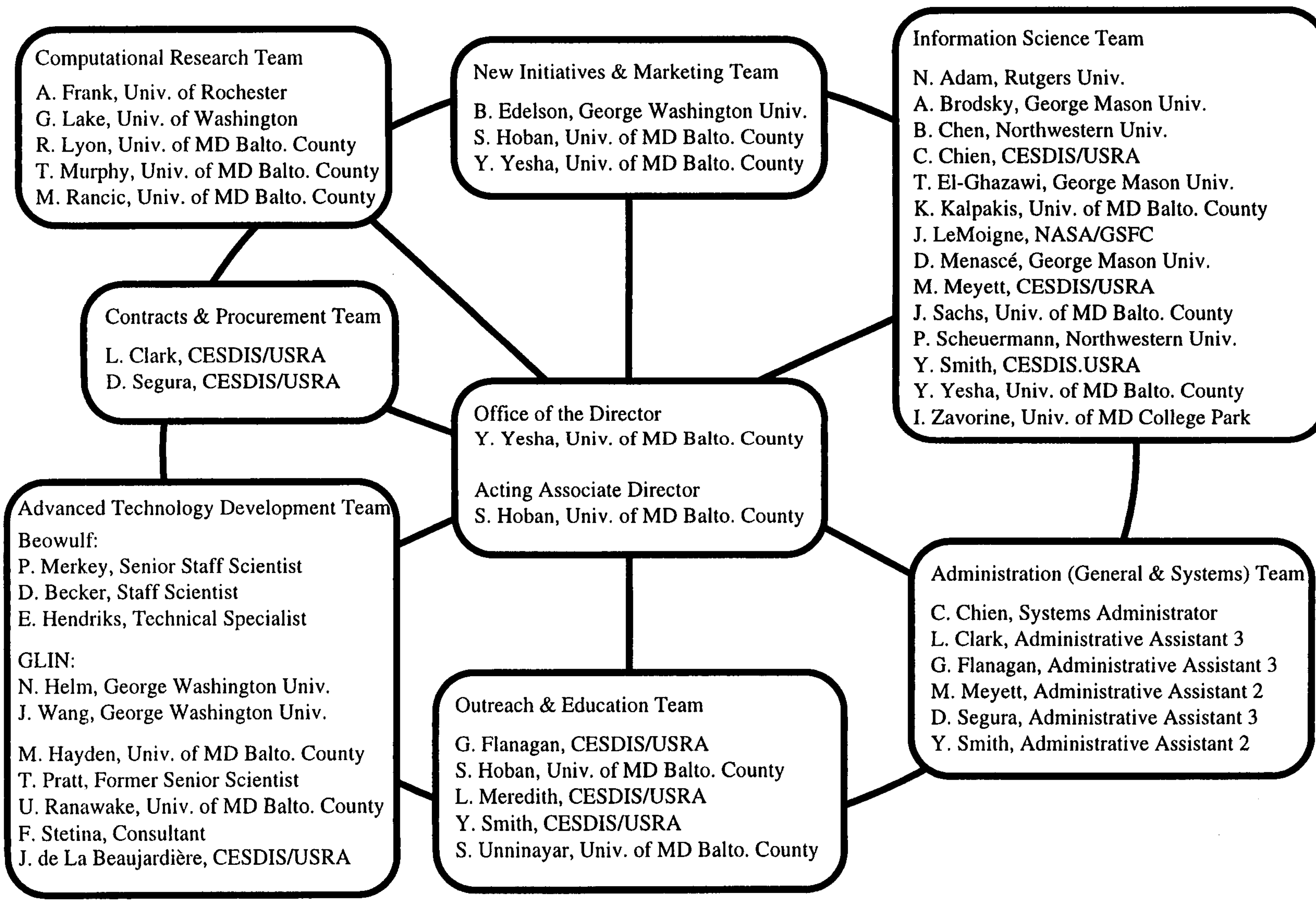
http://cesdis.gsfc.nasa.gov/

Contained in this web site are an overview of the CESDIS mission, special announcements, an explanation of the CESDIS organizational structure, and links to specific research projects and accomplishments.

The CESDIS home page is an active link to the heart of CESDIS activities. Feedback and comments are encouraged electronically to:

cas@cesdis.gsfc.nasa.gov

CESDIS ORGANIZATIONAL CHART



DIRECTOR

Dr. Yelena Yesha
(yelena@cesdis.usra.edu)

Dr. Yelena Yesha is a tenured full professor in the Department of Computer Science and Electrical Engineering at the University of Maryland Baltimore County (UMBC), hold a joint appointment with the University of Maryland's Institute for Advanced Computer Studies (UMIACS) in College Park, and serves as the CESDIS Director through a memorandum of understanding between the University of Maryland and USRA.

Dr. Yesha received a Bachelor of Science degree in computer science from York University in Toronto, Canada in 1984, and a Master of Science and Ph.D. in computer and information science from Ohio State University in 1986 and 1989 respectively. She is a Senior Member of the IEEE Society, and a member of the ACM and New York Academy of Science. Her research interests include distributed databases, distributed systems, and performance modeling. She has authored numerous papers and edited six books in these areas.

Prior to joining CESDIS in December 1994, Dr. Yesha was on leave from the University to serve as the Director of the Center for Applied Information Technology at the National Institute of Standards and Technology. The Center's mission was to advance the goals of the National Information Infrastructure by identifying, developing, and demonstrating critical new technologies and their applications which could be successfully commercialized by U. S. industry.

ACTIVITIES

I served as a general chair of IBM's International Workshop on Technological Challenges for Electronic Commerce. The workshop was attended by over 100 scientists from different countries.

Ms. Jeannie Behnke (Code 586) visited CESDIS and attended a meeting with Susan Hoban (CESDIS Acting Associate Director), Professor Kostas Kalpakis, and Joel Sachs (UMBC), to discuss the work on the new CESDIS task in the area of data warehousing.

I attended a seminar presentation titled "on quantum cryptography" given by Professor Sam Lomonaco (UMBC). His lecture was well attended and received. The topic of quantum cryptography generated substantial interest at Goddard.

I held a CESDIS staff meeting to discuss the upcoming CESDIS science council meeting. Also a topic of discussion at this meeting was the hiring of additional personnel.

Susan Hoban and I attended a Code 930 retreat at College Park, and gave a presentation on the current status of and future plans for CESDIS research.

I prepared for the Digital Earth conference that should take place in November and the workshop on computer simulation that is going to take place in January.

We had a visit by faculty and the Director of the Computational Science Institute from George Mason University. The purpose of the visit was to discuss the potential joint collaboration between CESDIS and GMU.

Susan Hoban (CESDIS Acting Associate Director) and I visited the USRA Headquarters in Columbia and held a meeting with Dr. Cummings, the USRA Executive Director, to discuss the future activities of CESDIS.

CESDIS hosted a seminar by Dr. Peter Norris from New Zealand.

Dr. Lemoigne left CESDIS and joined the civil service staff. We are planning to work on hiring her replacement. At this point CESDIS is advertising for a number of open positions.

I prepared for the Simulation Workshop that is scheduled to be held at CESDIS on Jan. 20-21, 1999.

I traveled to Cologne, Germany to attend a meeting of the GECOMMNET project, the International project that is tasked with the development of a global Masters program in Electronic Commerce.

I gave an invited lecture on data warehousing at INRIA France.

I attended the Digital Earth Workshop. The purpose of the workshop was to focus on the interagency program on Digital Earth. CESDIS is expected to play a major role in this program.

Ms. Irene Quarters, former VP of Cray Co., visited CESDIS and gave an invited lecture on the art of management of large software projects.

Professor Wolfson has visited CESDIS and worked with me and CESDIS scientists on research related to data mining.

I had a meeting with Dr. David Cummings, the executive Director of USRA, to discuss new initiatives.

I traveled to Toronto, Ontario in order to participate in CASCON98, IBM Annual Conference. I attended and co-chaired 8 workshops on Electronic Commerce at IBM CASCON98 Conference in Toronto, Canada.

Members of the Science Council came to CESDIS to conduct a review and hear presentations made by CESDIS scientists. This year we have 5 new members of the Science Council, so additional time was spent with them in order to familiarize them with the organization.

Traveled to Porto, Portugal to give a keynote speech at a major Portuguese conference on electronic commerce.

I attended a workshop on Information Technology at the National Institute of Standards and Technology, that was organized by the Advanced Technology Program.

Dr. Susan Hoban, Mr. Rick Lyon, Mr. Tim Murphy, and myself held a meeting at the USRA Washington office with Dr. Paul Coleman (USRA President). The topics of the discussion were new initiatives, and the impact of the optics group at CESDIS on space science.

I traveled to Maui Hawaii to attend the 32nd International Conference HICCS99. At HICCS99 I chaired a panel on the Technological Challenges in Electronic Commerce and attended a number of sessions where the new results on information technology were presented.

I visited Maui Supercomputing Center and met with number of scientists there, trying to identify possible area of collaboration between CESDIS and Maui Supercomputer Center.

I hosted an informal workshop on the role of remote sensing in containing and monitoring forest fires. The workshop was attended by Goddard Scientists and also, Scientists from IBM, University of Toronto and Australia. The joint follow up project activity is planned.

CESDIS held an International Workshop on Computer Simulation. The top scientists from all around the world attended and participated at the workshop. Mr. Al Diaz, the Goddard Center Director gave a keynote speech at the workshop.

I traveled to Livermore National Laboratory to give an invited lecture on the work in the area of Performance Modeling of Mass Storage Systems that I conducted at CESDIS.

My paper "Updating and Querying Databases that Track Mobile Units" has been recommended for publication in the Journal on Parallel and Distributed Databases special issue: Mobile Data Management and Applications.

Professor Kosaraju (John's Hopkins University) and I spent some time going over the current and future potential projects at CESDIS.

I attended the program committee meeting for the International Conference on Advances in Digital Libraries, that will be held in Baltimore on May 19-24, 1999. The program committee made the decision about the technical part of the program and social events that will be associated with this event.

I met with Dr. Rubens Medina (Law Library of Congress) to discuss the progress of the GLIN project and also the progress on the ELAS project.

I met with Dr. Rubens Medina (Law Librarian of Congress) to discuss the future of GLIN project. I made a significant progress in completing the research papers in the area of replication of databases.

I attended an International Conference on Data Engineering that took place in Sydney, Australia. I served as a member of Technical Program committee for this conference and also was a chair of the Session titled "Data management for Mass Storage Systems". Conference was attended by top researchers in the field and presenters delivered very impressive research results in the area of Information Technology. I held number of meeting with the researchers from academia and industry, and established new research collaborations for CESDIS.

I held a number of meetings with Dr. Edelson to discuss new initiatives in the area of Next Generation Internet.

I traveled to New York to attend a major convention in the area of Electronic Commerce. Additionally, I obtained new research results in the area of mobile electronic commerce.

I held a number of meeting with Dr. Cummings, (Executive Director of USRA) to discuss future plans for CESDIS.

I formally resigned as CESDIS Director, effective August 15, 1999. Dr. Susan Hoban (CESDIS Acting Associate Director) is expected to be designated as CESDIS Acting Director until the formal search commences to appointed the new permanent Director.

CONSULTANTS TO THE DIRECTOR

Task 1 is on the CESDIS contract (the general administrative task), and allows the Director to bring to CESDIS consultants who are not funded by specific task originators. CESDIS entered into agreements with the individuals reported upon in this section for the purpose of program development.

Maurice Aburdene
Bucknell University
Department of Electrical Engineering

Ian F. Akyildiz
Georgia Institute of Technology
Broadband and Wireless Networking Laboratory
(ian@ee.gatech.edu)

Burton I. Edelson
George Washington University
Department of Electrical Engineering and Computer Science
(edelson@seas.gwu.edu)

S. Rao Kosaraju
The Johns Hopkins University
Department of Computer Science
(kosaraju@cs.jhu.edu)

Richard Somerville
University of California, San Diego
Scripps Institution of Oceanography
(rsomerville@ucsd.edu)

Ouri Wolfson
University of Illinois at Chicago
Department of Electrical Engineering and Computer Science
wolfson@eecs.uic.edu

Control System Algorithms for Deformable Mirror Telescope

Maurice Aburdene
Bucknell University
Department of Electrical Engineering
(aburdene@bucknell.edu)

This report focuses on the options for control algorithms for an adaptive optics control system of the Next Generation Space Telescope. We begin this report by identifying possible optimization metrics. Due to the limited time available for investigating the total control system, emphasis has been devoted to a control system for the deformable mirror subsystem. It is assumed that a hierarchical control approach will be used in which "fine" control is performed by the deformable mirror subsystem.

Let the noise free control system representation for a deformable mirror telescope subsystem be as shown in Figure 1.

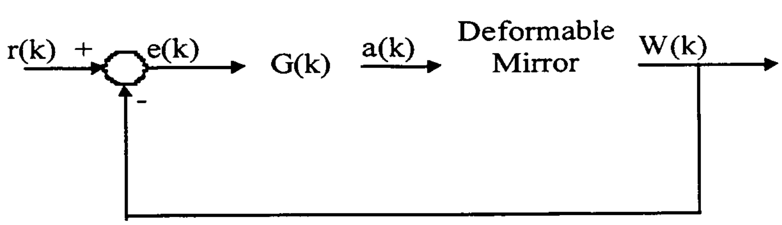


Figure 1: Telescope Control System

where

k is the discrete time instant,

Na is the number of actuators,

Ns is the number of sensed measurements

a(k) is an Na x 1 vector of computed actuator commands to the array of deformable mirrors,

W(k) is an Ns x 1 vector of phase values obtained from the phase retrieval algorithm,

R is an Ns x Na influence matrix for the deformable mirrors,

e(k) is an Ns x 1 wavefront error,

r(k) is an Ns x 1 reference vector, which is assumed to be equal to zero in our case, and

G(k) is an Na x Ns control matrix.

The computed actuator commands are sent to the actuators using a digital integrator [Furber and Jordan], [Corrigan, Furber, and Ramirez], [Wirth, Navetta, Looze, Hippler, Glindemann, and Hamilton] and [Grocott and Miller].

The deformable mirror relationship is given by

$$W(k) = R\, a(k)$$

Much more detailed block diagram representations of adaptive optics telescopes that include many sources of errors and actuator dynamics are presented in the papers by [Furber and Jordan], [Corrigan, Furber, and Ramirez], [Wirth, Navetta, Looze, Hippler, Glindemann, and Hamilton], [Dessenne, Madec, and Rousset] and [Lau, Breckenridge, Nerheim, and Redding], [Redding, Milman, and Loboda], [Redding and Breckenridge]. The diagrams will not be repeated here.

Performance Metrics or Performance Criteria

Optical Performance Metrics

a) Strehl ratio is "the ratio of the peak intensity at the focal point of the actual aberrated system to the peak intensity at the same point for a perfect, unaberrated system" [Redding, Milman, and Loboda, p. 91].

b) Encircled energy is "the amount of energy that fall into a particular region surrounding the focal point" [Redding, Milman, and Loboda, p. 91]

Control System Performance Metrics

a) Least squares: The figure of merit, performance criterion or objective function is to minimize

$$J = \sum_{1}^{N_S} e^2(k)$$

by proper choice of a(k)'s. This is the most common performance criterion.

b) Absolute value: The figure of merit, performance criterion or objective function is to minimize

$$J = \sum_{k=1}^{N_S} |e(k)|$$

by proper choice of a(k)'s.

c) Minimize maximum error: The figure of merit, performance criterion or objective function is to minimize

$$J = Min[Max\{|e(k)|\}]$$

Control Algorithms

Many control algorithms have been applied to the control of ground-based telescopes. The algorithms include: proportional integral derivative (PID), integral, linear quadratic regulator (LQR), linear quadratic Gaussian (LQG), fuzzy control, neural network, H2 control, H infinity control, adaptive neural net control, dynamic reconstructor control, adaptive control, nonlinear control, mixed technique control, and hierarchical control using some of the mentioned control techniques.

In this report we will assume that the transient response of the actuators is fast and we will focus on the steady state control.

1) Least squares without constraints: In this case our figure of merit or performance criterion or objective function is to minimize

$$J = \sum_{1}^{N_S} e^2(k)$$

by choosing the appropriate a(k)'s using the transfer function of both the deformable mirror and the measurement system, T. Here, we will assume that it is given by the transfer function of the deformable mirror, T=R. To obtain the optimal solution for the a(k)'s when Ns >> Na, we need to use the pseudoinverse of R, which is given by

$$a(k) = (R^T R)^{-1} R W(k)$$

Here, we note that in real systems, it is difficult to identify R and there is usually noise in the measurements of W(k). It should be clear that "better" control of the adaptive system will be obtained if we have a good estimate of T. There are efficient algorithms for finding the pseudoinverse without having to invert matrices and without forming

$$R^T R$$

using singular value decomposition techniques [Furber and Jordan].

2) Constrained least squares: In this case, our figure of merit is as shown in equation (?). However, we place limits on the values of a(k). In the case of our deformable mirror system, the constraints are given by

$$\alpha \le a(k) \le \beta$$

and for all neighboring actuators

$$-\gamma \le a_i(k) - a_j(k) \le \gamma$$

There are many algorithms that perform the constrained least squares algorithms and are included in the references [Dorn], [Murtagh and Saunders], and [Dixon] and on the Web page attached to this report.

Quadratic Programming

The quadratic programming problem with constraints is of the form

$$\text{Minimize} \quad a(k)^T B a(k) + C a(k)$$

where B is Na x Na matrix and C is 1x Na row vector.

If we now define our objective function J(a(k)) as in the least squares problem, then

$$(W(k) - Ra(k))^T (W(k) - Ra(k)) = W(k)^T W(k) - W(k)^T Ra(k) - (Ra(k))^T W(k) + (Ra(k))^T Ra(k)$$

Now the problem is to minimize

$$a(k)^T R^T Ra(k) - 2W(k)Ra(k) + W(k)^T W(k)$$

We note that the last term is independent of a(k) and therefore, the problem is in the quadratic programming form. As discussed earlier, we have the same constraints on a(k).

Noise Considerations

Assume that we now include noise in our phase error due to actuator input. Then phase error is given by:

$$W(k) = Ra(k) + d(k)$$

If d(k) is white noise and has a covariance matrix, X then by using the minimum variance criterion [Sorenson] and [Morrison] we obtain

$$a(k) = (R^T X^{-1} R)^{-1} R^T X^{-1} W(k)$$

This form is equivalent to the least squares form if

$$X = \sigma^2 I$$

where

I is the identity matrix and σ^2 is the variance of d(k).

Sequential Least Squares

Let us begin with the equation

$$W(k) = Ra(k) + d(k)$$

and assume that R is deterministic where

$$a(\hat{k}) = (R^T X^{-1} R)^{-1} R^T X^{-1} W(k) \qquad (1)$$

with error covariance matrix [Sorenson]

$$P = (R^T X^{-1} R)^{-1}$$

Following the approach of Sorenson (pp. 280-283), we can rewrite equation (1) in the following form with appropriate statistics for d(k) as:

$$a(\hat{k}) = a(k) + (R^T X^{-1} R)^{-1} R^T X^{-1} d(k)$$

and we can write

$$a(\hat{k}) = a(k-1) + K(k)[W(k) - Ra(\hat{k}-1)]$$

where

$$K(k) = P(k-1)R^T[(RP(k-1)R^T + X]^{-1} = P(k)R^T X^{-1}$$

$$P(k)\ldots E[(a(k) - a(\hat{k}))(a(k) - a(k))^{-T} = P(k-1) - K(k)RP(k-1) = (P^{-1}(k-1) + R^T X^{-1} R)^{-1}$$

$$P(0) = E[(a(k) - a(\hat{0}))(a(k) - a(\hat{0}))^T]$$

Stochastic Approximation

Let

$$W(k) = R[a(k)]a(k) + d(k)$$

Following the approach of Sorenson (pp. 288-307), we obtain

$$a(\hat{k}) = a(\hat{k}-1) + A(k)[W(k) - R(a(\hat{k}-1))]$$

The gain sequence A(k) should be chosen to allow for convergence. The suggested value for A(k) that is simple but non-optimal is given by

$$A(k) = R/|R(i,j)|^{**}2$$

Genetic Algorithm Approach

Genetic algorithms are stochastic in nature and have been applied to solve both constraint satisfaction and constrained optimization problems [In the reference by Chamber, Chapter 10 by Eiben, Raue, and Ruttkay] and includes C code], Fong, Cole, and Robertshaw], [Merino, Reyes, and Steidley, tutorial in nature] and [Berry and Linoff]. A fitness function is required and plays an important role in the convergence of genetic algorithms. One of the optimization criteria can serve as the fitness measure subject to the constraints described earlier. Eiben Raue, and Ruttkay applied genetic algorithms to a variety of problems and were encouraged by their results in constrained optimization problems.

Yim and Kyung have used a genetic algorithm and simulated annealing to minimize the track density and interconnection delay in a datapath area. They reported that a genetic algorithm combined with a simulated annealing algorithm is faster than simulated annealing alone. Their results indicate that simulated annealing produces better results than genetic algorithms, but had slower convergence initially.

Fong, Cole, and Robertshaw provided a comparison of various genetic algorithms in feedback controller design to minimize a quadratic performance criterion(LQR). They concluded that much more work is needed if the system model is not known or is changing and further investigation is needed to apply genetic algorithms to adapt to real-time control systems. They recommend a hybrid approach of genetic algorithms and fuzzy or neural control [Sandler, Barret, Palmer, Fugate, and Wild]. Hrycej presents an excellent overview of practical applications of neurocontrol. Schalkoff presents fundamental ideas of artificial neural networks(ANNs). Looney focuses on feedforward artificial neural networks that are well suited for decision making.

Suggestions For Further Study

1. Study the effects of using various performance measures presented to address the needs of the scientific community.
2. Identify the control limits of primary, secondary, and deformable mirror and determine the need for hierarchical control strategies.
3. Identify the transfer function of the optical system and develop a control system model of the telescope.
4. Obtain a good estimate of the influence function of the deformable mirror.
5. Develop a strategy for identifying variations in the influence function.
6. Combine off-line and on-line techniques for identification and control of the telescope subsystems.
7. Develop an on-line estimator of the influence function.
8. Concentrate initial effort on constrained least squares control methods and quadratic programming methods. Studies need to compare the speed and memory requirements of various algorithms.
9. Determine the applicability and efficiency of neural control strategies.
10. Determine the applicability and efficiency of genetic algorithms for optimization and control.
11. Determine the applicability and efficiency of simulated annealing for optimization and control.
12. Identify and investigate algorithms for control of unknown or time-varying influence functions.

13. Try suboptimal control strategies such as using unconstrained least squares and limiting actuator displacement to the boundries.

References

Bell, T. E. (1995). Electronics and the stars [telescope control]. *IEEE Spectrum, 32*(8), 16-24.

Berry, Michael J. A. and Linoff, Gordon (1998). *Data Mining Techniques: For Marketing and Customer Support.* John Wiley & Sons, Inc.

Bing Shen, and Stutzman, W. L. (1995). A scanning spherical tri-reflector antenna with a moving flat mirror. *Antennas and Propagation, IEEE Transactions, 43*(3), 270-276.

Chambers, Lance (1995). *Practical handbook of genetic algorithms: applications, 1*, CRC Press, Inc.

Chellabi, A., Stepanenko, Y., and Dost, S. (1997). Active shape control of a deformable mirror in an adaptive optics system. *Communications, Computers and Signal Processing, 10 Years PCRIM 1987-1997—Networking the Pacific Rim, 2*, 581-584.

Corrigan, William H., Furber, Mark E., Ramirez, Daniel F., and Cox, Charles D. Control software development for a large adaptive optical system. *SPIE, 2479*, 377-385.

Dessenne, Caroline, Madec, Pierre-Yves, and Rousset, Gerard (1998). Optimization of a predictive controller for closed-loop adaptive optics. *Applied Optics, 37*(21), 4623-4633.

Deval, A., Khan, A., and Akers, L.A. (1997). Locally adaptive active pixel sensors. In Circuits and Systems. ISCAS '97, *Proceedings of 1997 IEEE International Symposium, Vol. 1*, pp.685-688.

Dixon, L. C. W. (1972). *Nonlinear Optimization*, Crane, Russak, & Company, Inc.

Ellerbroek, B.L. and Plemmons, R.J. (1996). Computations in astro-imaging. In Image Processing, *Proceedings., International Conference, Sept. 16-19, 1996, Vol. 3*, pp. 113-116.

Fong, N.H. Benjamin, Cole, Daniel G., and Robertshaw, Harry H. A Comparison Study of Genetic Algorithms In Feedback Controller Design. *SPIE, 2443*, 727-738.

Furber, Mark E. and Jordan, David (1997). Optimal design of wavefront sensors for adaptive optical systems: part 1, controllability and observability analysis. *Opt. Eng. 36* (7), 1843-1855.

Furber, Mark E. and Jordan, David (1997). Optimal design of wavefront sensors for adaptive optical systems: part 2, optimization of sensor subaperture locations. *Opt. Eng. 36*(7), 1856-1871.

Grocott, Simon C.O. and Miller, David W. (1997). Dynamic Reconstruction and Multivariable Control for Force-Actuated, Thin Facesheet Adaptive Optics. *MIT Space Engineering Research Center*, SERC#3-97.

Gruneisen, M.T., Peters, K.W., Wilkes, J.M., and Peterson, P. (1996). Compensating primary mirror distortions in astronomical telescopes with real-time holography. *Lasers and Electro-Optics Society Annual Meeting, (LEOS 96.), IEEE, Vol. 1*, pp. 356-357.

Hrycej, Tomas (1997). *NEUROCONTROL:Towards an Industrial Control Methodology*, John Wiley & Sons, Inc.

Huang, J. and Castanon, D.A. (1997). Nonlinear estimation of misalignment errors in adaptive optics. In Decision and Control, *Proceedings of the 36th IEEE Conference, Vol. 1*, pp. 840-845.

Huang, J., Looze, D.P., Denis, N., and Castanon, D.A. (1995). Control designs for an adaptive optics system. In Decision and Control, *Proceedings of the 34th IEEE Conference, Vol. 4*, pp. 3753-3756.

Huang, J., Looze, D., Denis, N., Castanon, D. and Wirth, A. (1995). Dynamic modeling and identification of an adaptive optics system. In Control Applications, *Proceedings of the 4th IEEE Conference*, pp. 456-463.

Hyde, T. Tupper (1997). Vibration Isolation and Structural Control Technology for NGST. *NGST Technology Challenge*, 1-19 (NGST Webpage).

Jain, S. and Khorrami, F. (1997). Decentralized adaptive control of a class of large-scale interconnected nonlinear systems. *Automatic Control, IEEE Transactions, 42*(2), 136-154.

Junghsen Lieh, Wei Jie Li, and Kladitis, P. (1997). Intelligent vision-based control system. *Aerospace and Electronics Conference. (NAECON 1997), Proceedings of the 1997 National, Vol. 2*, pp. 792-797.

Kanev, Fedor, Lukin, Vladimir, and Lavrinova, Lidiya (1998). Analysis of algorithms for adaptive control of laser beams in a nonlinear medium, *Applied Optics, 37*(21), 4598-4603.

Knowles, Gareth J. Hybrid^2 Control Design for Intelligent Structures. *SPIE, 2443*, 864-872.

Landau, I.D., Lozano, R., and M'Saad, M. (1998). *Adaptive Control Systems*, Springer-Verlag.

Lau, Kenneth, Breckenridge, Bill, Norheim, Noble, and Redding, David (1992). Active Figure Maintenance Control using an Optical Truss Laser Metrology System for a Space-Based Far-IR Segmented Telescope. *SPIE, Controls of Optical Systems, 1696*, 61-82.

Looney, Carl, G. (1997). *Pattern Recognition Using Neural Networks: Theory and Algorithms for Engineers and Scientists*, Oxford University Press.

Lyon, Richard G., Hollis, J.M., Dorband, John E., and Murphy, Timothy P. (1998). Extrapolating HST Lessons to NGST. *Optics & Photonics News*, 35-39.

Mehmel, C., Caron, B., Marion, F., Mours, B., and Flaminio, R. (1997). Active control of suspensions for a gravitational wave detector. *Control Applications, Proceedings of the 1997 IEEE International Conference*, pp. 451-456.

Merino, Dennis I., Reyes, Edgar N., and Steidley, Carl (1998). Genetic Algorithms: Theory and Application. *Proceedings 1998 ASEE Conference, Session 2220.*

Morrison, Norman (1969). *Introduction to Sequential Smoothing and Prediction*, McGraw-Hill Book Company.

Murtagh, B.A., (1978). Large-Scale Linearly Constrained Optimization. *Mathematical Programming, 14*, 41-72.

Nelson, Jerry (1989). The Keck Telescope. *Scientist, 77*, 170-176.

Redding, David C. (1991). Optical Modeling for Dynamics and Control Analysis. *Journal of Guidance, Control and Dynamics, 14*(5), 1021-1032.

Redding, David C., Milman, Mark, and Loboda, Greg (1992). Linear Analysis of Opto-Mechanical Systems. *SPIE, Controls for Optical Systems, 1696*, 84-100.

Rifkin, A. (1997). Reengineering the Hubble space telescope control center system. *IEEE Internet Computing, 1*(3), 28-35.

Sandler, D. G., Barrett, T.K., Palmer, D.A., Fugate, R.Q. and Wild, W.J. (1991). Use of a neural network to control an adaptive optics system for an astronomical telescope. *Nature, 351*, 300-302.

Schalkoff, Robert J. (1997). *ARTIFICIAL NEURAL NETWORKS*, The McGraw-Hill Companies, Inc.

Seery, Bernard D. and Smith, Eric P. (1998). NASA's Next Generation Space Telescope: Visiting a Time When Galaxies Were Young. *Optics & Photonics News*, 29-34.

Sorenson, Harold W. (1980). *Parameter Estimation: Principles and Problems*, Marcel Dekker, Inc.

Taiwei Lu and Lerner, J. (1996). Spectroscopy and hybrid neural network analysis. *Proceedings of the IEEE, Volume 84*(6), pp. 895-905.

Tay, Teng-Tiow (1998). *High Performance Control*, Birkhauser Boston.

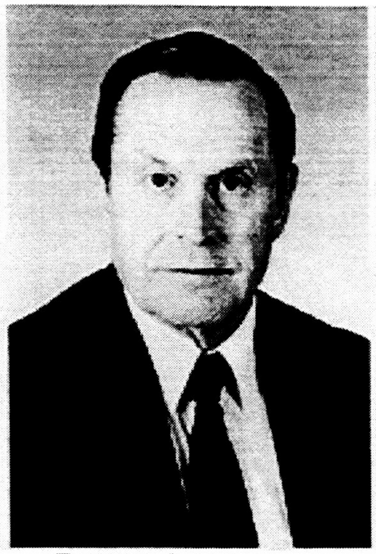
Trueblood, Mark, and Genet, Russell M. (1985). *Microcomputer control of telescopes*, Willmann-Bell, Inc.

Wirth, Allan, Navetta, Joseph, Looze, Douglas, Hippler, Stefan, Glindemann, Andreas, and Hamilton, Donald (1998). Real-time modal control implementation for adaptive optics. *Applied Optics, 37*(21), 4586-4597.

Yim, J.-S. and Kyung, C.-M. (1998). Datapath layout optimization using genetic algorithm and simulated annealing. *IEEE Proc.-Comput. Digit. Tech., 145*(2), 135-141.

Ian F. Akyildiz
Georgia Institute of Technology
Broadband and Wireless Networking Laboratory
(ian@ee.gatech.edu)

Dr. Akyildiz participated in the coordination of the organization of the workshop on the "Roles of Computer Science" to celebrate the 10th Anniversary of CESDIS. See pages 145-148.



Burton I. Edelson
George Washington University
Department of Electrical Engineering and Computer Science
(edelson@seas.gwu.edu)

Goals

Provide expertise to CESDIS in satellite communications and high performance networking; and plan and organize CESDIS cooperative projects with NASA, other US government agencies, U.S. industry, and, where appropriate, foreign research organizations.

Activities

1. Provided technical expertise in satellite communications to NASA GSFC. Led effort to get NASA and CESDIS involved in satellite communications and G-7 Information Society programs. Worked with Pat Gary (NASA GSFC) to arrange for and conduct several high data rate transmission tests involving satellite and fiber-optic links.

2. Worked with Pat Gary (NASA GSFC) to plan and develop the Testbed for Space and Terrestrial Interoperability (TSTI) to test the capability and develop procedures for satellite and optical fiber links to be inter-connected in high data rate networks. This testbed utilizes satellites and the ATDNet to develop transmission and networking procedures, standards, protocols, and equipment necessary to interconnect networks at data rates of 45, 155, and 622 Mb/s.

3. Worked with Pat Gary (NASA GSFC), Susan Hoban (NASA GSFC), Neil Helm (GWU), Eddie Hsu (JPL) and others on arranging a set of digital library experiments to connect U.S. data archives at the Library of Congress, National Library of Medicine, Department of Agriculture, and NASA GLOBE data center with corresponding data centers in Japan.

4. Continued work led by Milt Halem (NASA GSFC), and supported by Yelena Yesha (NASA GSFC), Susan Hoban (NASA GSFC), and others from GSFC and CESDIS to develop and expand the Global Legal Information Network (GLIN) with the Law Library of Congress. Co-authored GLIN system plan. Worked with Pat Gary (NASA GSFC) and Neil Helm (GWU) to

procure two Ku-band satcom terminals, one to be installed at NASA Goddard and the other at distant locations to perform GLIN system demonstrations.

5. Completed work on executive panel for survey of global satellite communications sponsored by NASA and NSF. Reviewed and edited report on "Global Satellite Communications Technology and Systems" (300+ page report published December 1998).

6. Worked with Sam Venneri and Ramon DePaula (NASA HQ) and NASA centers, to plan inter-center coordination and cooperation in satcom and high performance networking.

7. Hosted visit of Dr. Takashi Iida and Dr. Naoto Kadowaki of the Communications Research Lab, Ministry of Posts and Telecommunications, Japan, to consider plans for a cooperative program with NASA and U.S. industry to participate in the "Gigabit Satellite" program. Followed up with Alan Ladwig and Sam Venneri (NASA HQ) to generate a U.S. Government/industry position.

8. Visited NASA centers (GSFC, ARC, JPL, GRC and JSC) and NASA Institute for Advanced Concepts (NIAC) during the year to organize and coordinate R&D projects.

Conferences and Workshops

Japan-US Science Technology and Space Applications Program (JUSTSAP) workshop - Hawaii, November 9-13, 1998

USRA-NIAC Science Symposium, Washington DC, March 25-26, 1999

IEEE Advances in Digital Libraries (ADL-99) conference, Baltimore MD, May 19, 1999

Publications

Edelson, B. I. and Helm, N. R. (1997). High Data Rate Satellite Communications: Interoperability Issues (IAF-97-M.1.09). *48th International Astronautical Congress*, Turin, Italy.

Helm, N. R. and Edelson, B.I. (1997). Space Technologies and Systems for Disaster Mitigation, (IAF-97.C.2.01). *48th International Astronautical Congress*, Turin, Italy.

Pelton, J.N. et al (1998). Global Satellite Communications Technology and Systems. *International Technology Research Institute*, (Report of study sponsored by NASA and NSF, 330 pages).

A Data Organization for Storing and Searching Large Data Sets

S. Rao Kosaraju
The Johns Hopkins University
Department of Computer Science
(kosaraju@cs.jhu.edu)

The purpose of this research has been to develop an efficient data structure for storing and searching large data sets. It is assumed that the data set is stored in secondary storage, and hence the goal is to reduce the number of accesses to the storage.

In the simplest model we assume that the data consists of a large number of words, each word being a string of characters. The queries consist of searching for a given word (*member*), inserting a new word (*insert*), or deleting an existing word (*delete*). Our goal is to design a simple data structure that permits efficient execution of the queries. We have developed a data structure that exploits the advantages of two existing tree-based data structures, standard tries and balanced search trees, while avoiding their disadvantages.

The standard *trie* is constructed by repeatedly partitioning the set of words into subsets based on the first discriminating character. Figure 1 shows the trie for the set S = {*aaaa, aaaba, aabba, aabbb, aba, abbaaa, abbaab, abbab, abbbaa, abbbab, bbbba, bbbbba*}. Note that at the root the set is split into two subsets since the first character can be *a* or *b*. All the words that start with *a* (respectively *b*) have *a* (respectively *bbbb*) as a common prefix; hence the label of the left (respectively right) edge is a (respectively *bbbb*). To search for membership of a word, say, *abbbb*, starting from the root we follow the edges labeled *a, b, b, ba* and then decide that the given word is not in the set S.

The balanced search tree is constructed by repeatedly partitioning the set of words, after sorting lexi-cographically, into two equal sized subsets. Figure 2 shows this structure for the set S. Note that at every node the corresponding set of words is split into two equal sized subsets. Searching for membership of a given word is complicated and requires exploring multiple paths from the root.

The standard trie construction can result in very deep trees, while the balanced search tree guarantees a depth of log n, where n is the number of words. However, member searches in balanced search trees is very inefficient. Our model, denoted *balanced trie*, guarantees log n depth while preserving the advantages of the trie in performing member queries.

In a balanced trie, the symbol-based splits and the balanced splits alternate. Figure 3 shows this structure for the set S. At the root the split is based on the first symbol. Note that the left (respectively right) child of the root splits the 10 (respectively 2) words that start with *a* (respectively *b*) into two blocks each having 5 (respectively 1) words.

Member searches in balanced tries are extremely easy and efficient.

Planned Work

So far we have been able to handle the case when the data set S is given at the beginning. For this static case, we have designed an efficient algorithm for constructing a balanced trie. We plan to extend the approach when new words can be inserted and existing words can be deleted.

A second objective is to implement the algorithms and study the performance for large data sets.

Another goal is to apply the algorithms to word-based data compression algorithms. In these algorithms another search model is more appropriate. For the compression problems, the data set is a single very long word, T. Given a word P, our problem is to search whether P is a subword of (occurs in) T. A classic approach to this problem requires the construction of a trie for all the suffixes of T. This data structure, known as the suffix tree, can be as deep as the number of characters of T. It is easily seen a balanced suffix tree, based on our balanced trie, guarantees a depth of order log n, where n is the number of characters in T. We plan to design an efficient algorithm for the construction of a balanced suffix tree. We will apply the resulting algorithm for dictionary-based approaches to data compression.

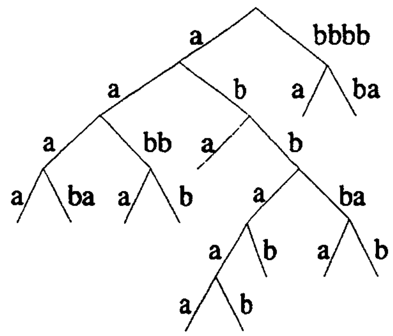


Figure 1: Trie for set *S*

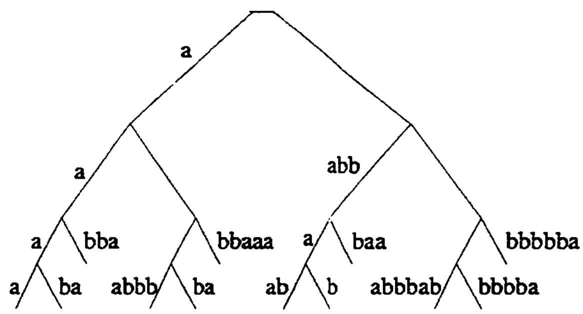


Figure 2: Balanced Search Tree

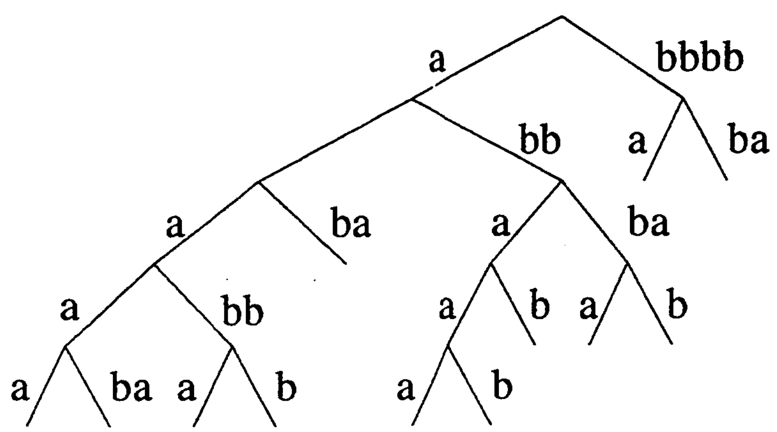


Figure 3: Balanced Trie

References

Chang, W.L. and Lawler, E. L. (1990). Approximate string matching in sublinear expected time. *Proc. of g1t IEEE Symp. on Foundations of Computer Science*, pp. 116-124.

Chen, M.T. and Seiferas, J. (9185). Efficient and elegant subword tree construction. In A. Apostolico and Z. Galil, (editors), *Combinatorial Algorithms on Words*, 97-107. Springer-Verlag, New York.

Knuth, D.E. (1973). *The Art of Computer Programming, Volume 3 (Sorting and Searching)*, Addison-Wesley Publishing Co., Reading, Mass.

Kosaraju, S. R. (1994). Realtime pattern matching and quasi-real-time construction of suffix trees. *Proc. of 26th Annual ACM Symp. on Theory of Computing*, pp. 310-316.

McCreight, E. M. (1976). A space-economical suffix tree construction algorithm. *J. of ACM*, 262-272.

Salomon, D. (1998). *Data Compression - The Complete Reference*, Springer-Verlag, New York.

Weiner, P. (1973). Linear pattern matching algorithms. *Proc. of 14th IEEE Symp. on Suitchzng Automata Theory*, pp. 1-11.

Richard Somerville
University of California, San Diego
Scripps Institution of Oceanography
(rsomerville@ucsd.edu)

1. Building in-house CESDIS atmospheric science capability

In continuing discussions with Dr. Halem (NASA GSFC), I am working toward the long-term goal of building an in-house research capability in atmospheric science, comparable to the existing ones in space science and computer science. This goal includes the development of a strong collaborative research program with university scientists, but it also will require hiring to increase the in-house expertise. Most recently, I have helped to recruit a suitable Ph. D. atmospheric scientist to join the group at Goddard through CESDIS. He is Dr. Peter Norris.

2. Analysis of scientific purposes of future geostationary missions.

The following areas have been analyzed

A. MESOSCALE RESEARCH. Geostationary orbit allows fine time resolution and hence is essential for observing mesoscale phenomena, because these have too short time scales to be observable from polar orbit. These phenomena are not only interesting and important in themselves, but they also have many crucial climate implications. For example, tropical mesoscale cloud clusters are energetically important to the tropical general circulation, to the driving of the upward branch of the Hadley cell, to supplying water vapor (and heat and momentum) from the boundary layer to the upper tropical troposphere, etc. And they are important modulators of both solar and terrestrial radiation, which is likely to play a role in El Niño and related phenomena; e. g., convective clouds over high sea surface temperatures (SSTs) can decrease surface insolation and hence reduce SST. Some mesoscale phenomena, notably hurricanes, are among the most poorly understood and poorly predictable severe weather events, and these too have climate implications. Will a greenhouse-enhanced climate produce higher SSTs or larger regions of sufficiently high SST (around 28C is the threshold) for hurricanes to form? Will hurricane seasons be longer, or more widespread? Is Hurricane Mitch a foretaste of monster hurricanes of the future? We do not know, and research on hurricanes and their dependence on the climate regime in which they occur will require geostationary satellite observations. There are many other tropical and subtropical mesoscale examples that are both intrinsically important and have climate implications - the Indian monsoon onset involves a whole class of such phenomena. It will take creative use of geostationary platforms to make headway on these research issues.

B. CLOUD-RADIATION INTERACTIONS AND EARTH RADIATION BUDGET RESEARCH. Clouds have small spatial and temporal scales (kilometers and less, minutes), they are responsible for most of the planetary albedo, and they contribute powerfully to infrared trapping (the greenhouse effect). Therefore, it is critical that we not only monitor these cloud-dependent fields (cloud extent, water content, particle size, phase, radiative properties, radiation budget components, hydrologic cycle components, etc.), but that we do research on understanding them and ultimately incorporating them with sufficient realism in GCMs, i. e., parameterize their ensemble effects. At the same time, we need to monitor quantities such as the top-of-atmosphere and surface radiation budgets on the same space and time scales as the clouds, which control these budgets to a large extent. This requirement arises not just from the need to develop climatologies of these quantities, which involves averaging that sometimes makes geostationary resolution less essential, but also to increase basic understanding of the physics responsible for the variability. This kind of work is key to making progress on the number one priority on everybody's list for reducing the uncertainty in GCM estimates of climate sensitivity to greenhouse gases. Once the global uncertainty is reduced, there will still be a huge amount of work to be done in attacking the problem of regional and transient climate change, involving many different cloud types, dependence on seasons and synoptic regime, etc. This will keep geostationary birds flying for decades!

C. SYNERGISTIC RESEARCH AND CONTRIBUTIONS TO FIELD EXPERMENTS. The history of recent field experiments in the tropics provides an illustration of the role of geostationary observations, among recent experiments, TOGA-COARE and CEPEX fall into this category. The INDOEX field phases were judged sufficiently dependent on geostationary data that one satellite was moved from its usual longitude so that it could observe the INDOEX region. Some of the uses of the data are quite ordinary - helping to plan aircraft observations in real time, for example. But often the value of the geostationary measurements is that they are used in combination with other data from different in situ and satellite platforms. This is happening at the ARM site in Oklahoma, for example. Sometimes NASA gets caught up in the trap of mission-think, in that each mission gets planned and justified by the science problems that it might "solve," but in reality the greatest use of any observational data in this field, and certainly of geostationary satellite data, is often that it contributes an essential piece of the puzzle when combined with other data, and of course with models, not when it is used alone.

3. Planning and analysis for incorporation of improved cloud-radiation parameterizations in the Goddard modification of the NCEP Eta model for limited domains appropriate to the Triana mission

My main effort during this period has been aimed at developing, improving, testing and validating parameterizations of cloud-radiation interactions for climate models. These parameterizations are algorithms which express the influence of cloud-radiation effects on the climate system, an aspect of climate physics which is as yet poorly understood. Until recently, major global climate models all used simple empirical representations of clouds, based on arbitrary functions of relative humidity, tuned crudely to reproduce satellite measurements of the Earth's radiation budget. Using data from both ARM and TOGA-COARE, I have been able to show convincingly the shortcomings of these traditional diagnostic cloud schemes based on relative humidity, and also to demonstrate the potential gains to be found in using prognostic schemes based on predicted cloud water budgets and radiative properties derived from them. For additional details, see two of my recent papers:

- Lee, W.-H., Iacobellis, S. F. and Somerville, R. C. J. (1997). Cloud-radiation forcings and feedbacks: General circulation model tests and observational validation. *Journal of Climate, 10*, 2479-2496.

- Lubin, D., Chen, B., Bromwich, D. H., Somerville, R. C. J., Lee, W.-H., and Hines, K. M. (1998). The impact of Antarctic cloud radiative properties on a GCM climate simulation. *Journal of Climate, 11*, 447-462.

We have used a single-column model (SCM) diagnostically to evaluate cloud-radiation parameterizations against observations from the Atmospheric Radiation Measurement (ARM) Program. Cloud-radiation parameterizations display a strong sensitivity to vertical resolution in the SCM, and vertical resolutions typically used in global models are far from convergence. We have tested newly developed advanced radiation parameterizations in addition to radiation routines used in current general circulation models. We find that schemes with explicit cloud water budgets and interactive radiative properties are potentially capable of matching observational data closely. In our SCM, using an interactive cloud droplet radius decreases the cloud optical thickness and cloud infrared emittance of high clouds, which acts to increase both the downwelling surface shortwave flux and the outgoing longwave radiation. However, it is difficult to evaluate the realism of the vertical distribution of model-produced cloud extinction, cloud emittance, cloud liquid water content and effective cloud droplet radius until high-quality observations of these quantities become more widely available. We also find that in the SCM, cloud parameterizations often underestimate the observed cloud amount, and that ARM observations indicate the presence of clouds while the corresponding maximum relative humidity is less than 80%. This implies that the underlying concept of a critical gridpoint relative humidity of about 80% for cloud formation, as used in many cloud parameterizations, may need to be re-examined.

4. Climatic role of biomass burning aerosols

I have also investigated, with NASA support, the climatic role of aerosols from biomass burning. Among the interesting results, we find that these aerosols backscatter sunlight in cloudy conditions with an efficiency of 0.53, which is greater than that reported for sulfate aerosols. For details, see my paper:

- Iacobellis. S. F., Frouin, R. and Somerville, R. C. J. (1999). Direct climate forcing by biomass-burning aerosols: Impact of correlations between controlling variables. *Journal of Geophysical Research, 104*(D10), 12,031-12,045.

Ouri Wolfson
University of Illinois at Chicago
Department of Electrical Engineering and Computer Science
wolfson@eecs.uic.edu

Report

Consider a data warehouse that represents information about moving objects and their location. For example, for a data warehouse representing the current location of objects in a battlefield a typical query may be: retrieve the friendly helicopters that are in a given region, or, retrieve the friendly helicopters that are expected to enter the region within the next 10 minutes. The queries may originate from the moving objects, or from stationary users. We will refer to the above applications as MOtion-Database (MOD) applications or moving-objects-database applications.

In the military, MOD applications arise in the context of the digital battlefield, and in the civilian industry they arise in transportation systems. For example, Omnitracs developed by Qualcomm is a commercial system used by the transportation industry, which enables MOD functionality. It provides location management by connecting vehicles (e.g., trucks), via satellites, to company databases.

Currently, MOD applications are being developed in an ad hoc fashion. Data warehousing and Database Management System (DBMS) technology provides a potential foundation for MOD applications, however, DBMS's are currently not used for this purpose. The reason is that there is a critical set of capabilities that have to be integrated, adapted, and built on top of existing DBMS's in order to support moving objects databases. The added capabilities include, among other things, support for spatial and temporal information, support for rapidly changing real time data, new indexing methods, and imprecision management.

In this project we addressed the imprecision problem. The location of a moving object is inherently imprecise because, regardless of the policy used to update the database location of a moving object (i.e. the object-location stored in the database), the database location cannot always be identical to the actual location of the object. There may be several location update policies, for example, the location is updated every **x** time units. In this project we addressed threshold-policies, i.e. policies that update the database whenever the distance between the actual location of a moving object **m** and its database location exceeds a given threshold **h**, say 1 mile. This means that the DBMS will answer a query "what is the current location of **m**?" by an answer **A**: "the current location is **(x,y)** with a deviation of at most 1 mile".

One of the main issues addressed in this project was how to determine the update threshold **h** in such policies. This threshold determines the location imprecision, which encompasses two related but different concepts, namely deviation and uncertainty. The deviation of a moving object **m** at a particular point in time **t** is the distance between **m**'s actual location at time **t**, and its database location at time **t**. For the answer **A** above, the deviation is the distance between the actual location of **m** and **(x,y)**. On the other hand, the uncertainty of a moving object **m** at a particular point in time **t** is the size of the area in which the object can possibly be. For the answer **A** above, the uncertainty is the area of a circle with radius 1 mile. The deviation has a cost (or penalty) in terms of incorrect decision making, and so does the uncertainty. The deviation (resp. uncertainty) cost is proportional to the size of the deviation (resp. uncertainty). The ratio between the costs of an uncertainty unit and a deviation unit depends on the interpretation of an answer such as **A** above.

In MOD applications the database updates are usually generated by the moving objects themselves. Each moving object is equipped with a Geographic Positioning System (GPS), and it updates its database location using a wireless network (e.g ARDIS, RAM Mobile Data Co., IRIDIUM, etc.). This introduces a third information cost component, namely communication. For example, RAM Mobile Data Co. charges a minimum of 4 cents per message, with the exact cost depending on the size of the message. Furthermore, there is a trade-off between communication and imprecision in the sense that the higher the communication cost the lower the imprecision and vice versa. In this paper we propose a model of the information cost in moving objects databases, which captures imprecision and communication. The trade-off is captured in the model by the relative costs of an uncertainty unit, a deviation unit, and a communication unit.

Based on these cost-based trade-off principles we devised and analyzed several threshold-policies.

An additional contribution of this project is a probabilistic model and an algorithm for query processing in motion data warehouses. In our model the location of the moving object is a random variable, and at any point in time the database location and the uncertainty are used to determine a density function for this variable. Based on this model we developed an algorithm that processes range queries such as Q='retrieve the moving objects that are currently inside a given region R'. The answer to Q is a set of objects, each of which is associated with the probability that currently the object is inside R.

Publications

Tayeb, J., Ulusoy, O., Wolfson, O. (1998). A Quadtree Based Dynamic Attribute Indexing Method. *Computer Journal, 41*(3), 185-200.

Sistla, P., Wolfson, O., Yesha, Y. and Sloan, R. (1998). Towards a Theory of Cost Management for Digital Libraries and Electronic Commerce. *ACM Transactions on Database Systems (TODS), 23*(4), 411-452.

Wolfson, O., Sistla, P., Chamberlain, S. and Yesha, Y. (1999). Updating and Querying Databases that Track Mobile Units, invited paper, special issue of the *Distributed and Parallel Databases Journal (DAPD) on Mobile Data Management and Applications, 7*(3), 1999, Kluwer Academic Publishers, 257-288. (See \verb|http://www.cecs.missouri.edu/~joshi/dapdcall.html|).

Dao, S., Shek, E., Vellaikal, A., Muntz, R., Zhang, L., Potkonjak, M. and Wolfson, O. (1999). Semantic Multicast: Intelligently Sharing Collaborative Sessions, accepted to *ACM Computing Surveys*, to appear in March 1999.

Wolfson, O., Jiang, L., Sistla, P., Chamberlain, S., Rishe, N. and Deng, M. (1999). Databases for Tracking Mobile Units in Real Time. *Lecture Notes in Computer Science, number 1540, Proceedings of the 7th International Conference on Database Theory (ICDT)*, Jerusalem, Israel, pp. 169-186, Springer-Verlag.

Wolfson, O., Chamberlain, S., Sistla, P., Xu, B., and Zhou, J. (1999). DOMINO: Databases fOr MovINg Objects tracking. *Proceedings of the ACM-SIGMOD 1999, International Conference on Management of Data*, Philadelphia, PA, pp. 547-549.

Rishe, N., Naboulsi, K., Wolfson, O. and Ehlmann, B. (1999). An Efficient Web-based Semantic SQL Query Generator. *Proceedings of the 19th IEEE International Conference on Distributed Computing Systems Workshop*, Austin, TX, pp. 23-30.

Wolfson, O., Sistla, P., Xu, B., Zhou, J. and Chamberlain, S. (1999). Tracking Moving Objects Using Database Technology in DOMINO, accepted, to appear in the *Proceedings of NGITS'99, The Fourth Workshop on Next Generation Information Technologies and Systems*, Zikhron-Yaakov, Israel.

Wolfson, O., Lelescu, A. and Xu, B. (1999). Retrieval of Collaborative Work from Multimedia Databases Using Relevance Feedback, accepted to appear in the *Proceedings of Spire'99, Symposium on String Processing and Information Retrieval*, Cancun, Mexico, IEEE CS Press.

Sistla, P., Wolfson, O., Chamberlain, S. and Dao, S. (1998). Querying the Uncertain Position of Moving Objects, (invited) *Temporal Databases: Research and Practice,* Etzion, O., Jajodia, S., Sripada, S. (eds.), Springer-Verlag, *Lecture Notes in Computer Science number 1399*, pp. 310-337.

Rishe, Naphtali, Naboulsi, Khaled and Wolfson, Ouri (1999). Report Generators, *John Wiley and Sons Encyclopedia of Electrical and Electronics Engineering, Vol. 18*, 500-513.

INFORMATION SCIENCE TEAM

Nabil R. Adam
Rutgers University
Center for Information Management, Integration, and Connectivity (CIMIC)
(adam@adam.rutgers.edu)

Alexander Brodsky
George Mason University
Department of Information and Software Engineering (ISE)
(brodsky@gmu.edu)

Bin Chen
Northwestern University
Department of Electrical and Computer Engineering
(bchen@ece.nwu.edu)

Konstantinos Kalpakis
University of Maryland Baltimore County
Department of Computer Science and Electrical Engineering
(kalpakis@cs.umbc.edu)

Daniel A. Menascé
George Mason University
Department of Computer Science
(menasce@cs.gmu.edu)

Tarek El-Ghazawi
George Mason University
(Tarek@gmu.edu)

Jacqueline Le Moigne
NASA/GSFC
(lemoigne@backserv.gsfc.nasa.gov)

Ilya Zavorine
University of Maryland, College Park
Department of Mathematics
(iaz@cs.umd.edu)

Comparative Study of Commercial-of-the-Shelf-Client Query Tools to Support the CCS Data Warehouse

Nabil R. Adam
Rutgers University
Center for Information Management, Integration, and Connectivity (CIMIC)
(adam@adam.rutgers.edu)

Abstract

Data warehousing technique has been adopted by commercial and research communities to provide fast access to data originating from distributed, possibly heterogeneous, information sources. Data from information sources are integrated, sometimes summarized, and later stored in the data warehouse. The Hubble Space Telescope (HST) Vision 2000 Project has adopted data warehousing technology to provide fast access to engineering telemetry data and orbital events to HST instrument engineers and operators. While the data warehouse provides a single repository for the engineering telemetry data, the query tool provides end-users with the ability to access and analyze data stored in the data warehouse. Currently, queries to HST data warehouse are executed via customized queries or software specific data manipulation language (i.e. RISQL). For ad-hoc queries, the appropriate queries need to be prepared. Developing and supporting ad-hoc queries is inefficient. An alternative is to use commercially available data querying tools for parameterized and ad-hoc queries. In this case, the effort to develop customized queries and tools can be reduced considerably. This report presents a comparative study of commercially available query tools that would be suitable for HST users to use as an interface to the Control Center System (CCS) Data Warehouse.

Introduction

Both commercial and research communities have adopted the data warehousing techniques to provide fast access to integrated data originating from distributed, possibly heterogeneous, information sources. A data warehouse is simply a repository of data that have been extracted from multiple information sources, integrated, possibly summarized, and replicated. In order to access integrated data or to perform data analysis, end-users no longer need to access raw data at the underlying information sources. Data and their summaries are pre-prepared at the data warehouse and fast access to integrated data is supported. However, since the state of the data warehouse must be kept consistent with the state of the underlying information sources, the data warehouse needs to be updated (refresh) as soon as possible.

One of NASA's early goals for the Vision 2000 project was to have "All data on line and immediately available for operational use," a goal unattainable with classic technology. The Hubble Space Telescope (HST) Vision 2000 Project, in conjunction with the Space Telescope Science Institute has implemented the data warehousing approach to archiving and retrieving critical HST engineering telemetry data and orbital events. The HST team has selected a commercial data warehouse, Red Brick, as a single repository for their integrated HST data.

In this project, we focus on the querying-tool aspect of the data warehouse. At the present time, queries to the CCS Data Warehouse are submitted via CCS GUI, which provides HST users with customized queries. For ad-hoc queries, which are not pre-specified, the CCS GUI needs to be appended with the new customized query or a new RISQL command needs to be written. Expertise on writing RISQL query is required.

There is a need to be able to use query tools that can perform ad-hoc and parameterized queries, which are not pre-defined in the CCS GUI. Such query tools can potentially help in reducing the effort of developing new customized queries and writing new RISQL queries. Therefore, the goal of our project is to identify commercially available query tools that can be used by HST users to perform ad-hoc and parameterized queries to the CCS Data Warehouse. Comparison between the available commercial query tools is required to determine which tools are best suited to HST users' needs.

In the course of our discussion we detail a description of the environment on which the query tools need to be run. It discusses the design of the CCS Data Warehouse, some sample queries end-users normally post to the CCS Data Warehouse, and the client access to the data warehouse. We also present the criteria that are used to evaluate the performance and features of selected commercial query tools. The commercial query tools that have been selected for evaluations will also be discussed. Detailed evaluation of each commercial query tool is presented. For each query tool, the results of general evaluation, HST-user-query related evaluations, and additional evaluation criteria are provided. At the end, we present our recommendations.

Scope of Project

The purpose of this project is to conduct a survey of commercial client parameterized and ad-hoc query tools, which would be appropriate for use by users of the CCS Data Warehouse. Details of the scope of the research project follow:

- Conduct a Survey of Commercial-Off-The-Shelf (COTs) client query tools to support parameterized and ad-hoc queries to the CCS data warehouse.

- Develop prototypes using selected COTs candidate tools. For this task the investigative team is expected to develop prototypes using the selected client query products.

- Prototype Installation. Working jointly with CCS personnel, the investigation team is to participate in installation of the selected prototype in the CCS development environment.

The Environment

The CCS Data Warehouse mainly consists of the data loading processor, the data warehouse itself, and the CCS GUI and RISQL data manipulation language. A commercial data warehouse, Red Brick, was selected to house spacecraft telemetry and orbital events. The nightly load of telemetry data into the HST data warehouse takes approximately 15 minutes. A CCS data warehouse "preloader" provides a filtering process prior to the load. This takes about two and a half hours. As a result, the loading strategy makes telemetry data available from the data warehouse to any user within 24-30 hours of receipt by the HST ground system. Should a user need to query the most recently acquired data prior to this period, they can also access the all points archive via the CCS GUI.

At the present time the Data Warehouse will support the following queries from the CCS GUI interface for either changes only or averaged mnemonics: values within a time period for a set of given mnemonics; values for a given mnemonic within a time period when certain constraints are met; values for a list of mnemonics within a time period when constraints on one mnemonic are met; values for a list of mnemonics within a time period sampled at a selected time interval. Results are provided either as ASCII files to be viewed by the user or are packaged in FEP Output Format to allow processing by the CCS Analysis subsystem. Additional queries are planned for future releases. These queries will be based on a generic output format, which is currently being defined by the CCS project.

In this way the development team hopes to provide a simple format that is easily accommodated by a variety of COTS products used as either front-end or analysis tools. Additional queries planned for future releases include: Queries which provide qualitative information, for example: counts over time for a given set of conditions, whether specified conditions were met over a period, during what times specific conditions were met; and event queries against other information types to be possibly warehoused in the future (e.g.,. trend data).

Different Database Designs

Since Red Brick follows a relational approach to implementing a data warehouse, other approaches such as Multi-Dimensional databases are not discussed in this report. In designing a data warehouse implemented through a relational database approach, there are three approaches the warehouse administrator can use; Star, Snowflake or Fact-constellation schema. A star schema consists of a fact table and a set of dimension tables. All of the dimension tables are directly connected to the fact table. All the dimension tables are de-normalized.

End-User Queries

The queries submitted by the end-users include ad-hoc, simple and complex queries. Currently, queries to the data warehouse are executed via the CCS GUI. This GUI provides HST users with "canned" customizable queries. Queries can also be submitted directly to the CCS Data Warehouse through a command line interface to RISQL. Results are provided either as ASCII files to be viewed by the user or are packaged in Front End Processor (a specific CCS telemetry) output format to allow processing by the CCS Analysis subsystem. Additional queries are planned for future releases. These queries will be based on a generic output format, which is currently being defined by the CCS project. In this way the development team hopes to provide a simple format that is easily accommodated by a variety of COTS products used as either front-end or analysis tools.

Connecting to CCS Data Warehouse

In order to evaluate the selected commercial query tools, the actual HST data at CCS Data Warehouse are used. The query tools must make successful connection to the CCS Data Warehouse for that purpose. Currently, all the selected tools were installed onto machines at CIMIC - Rutgers University. A remote connection is made between every tool maintained at CIMIC and CCS Data Warehouse. The remote connection is made possible through the use of proprietary ODBC software provided by Red Brick. The platforms that are used to evaluate the tools are minimum configuration (i.e. Windows 95 and NT connecting to CCS RedBrick Warehouse through RedBrick Client ODBC).

Evaluation Criteria

In order to evaluate the selected commercial query tools, we have devised thorough evaluation criteria. The criteria are:

1. Support for the different warehouse designs
2. Support for HST-related queries
3. Overall features

Support for the Different Warehouse Designs

The first criteria being evaluated is the tool capability to support for the different warehouse designs: Star, Snowflake and Fact-constellation. For queries that require parameters be specified on dimensions tables directly related to the fact table, support for star schema design is required. For normalized dimension tables, snowflake design needs to be supported. The query tools should be capable of accepting end-users parameters on dimension tables not directly connected to the fact table. In this case, the user parameters on sub-dimensions need to be rolled-up to a higher dimension, which directly connected to the fact table. However, if many of the queries require drill-across the different fact tables, fact-constellation design should be supported. In this case, the query tool should be able to have the parameters be assigned not only to dimension tables, but also to the fact tables.

Support for HST-related queries

Once the query tools have been evaluated based on support for the different warehouse designs, they are further tested based on the specific needs of users of CCS Data Warehouse. We have used six different sample queries presented as the basis of queries of users of CCS Data Warehouse. For each query, the query tool is evaluated. If the query tool can perform the query, ease of use is measured. If the tool cannot perform the specified query, the reason is presented.

For Example:

```
Select mnem_friendly, start_time, stop_time, min_eu, max_eu, avg_eu
   From   average_tlm_1998 a, mnemonics m
   Where
        m.mnem_tag_id   = a.mnem_tag_id and m.mnem_friendly = 'F2SSCEA'
        and avg_eu  > 0 and start_date = '9/17/98' and start_time     <=
'04:00:00'
        and stop_time      = '03:50:00'

        OR

Select distinct t1.mnem_friendly, t1.start_time, t1.stop_time, t1.eu_value
   From
        (select mnem_friendly, start_time, stop_time, eu_value
         from  changeonly_tlm_1998 c, mnemonics m, discrete_codes d
         where  m.mnem_tag_id = c.mnem_tag_id and d.discrete_code_id =
c.discrete_code_id
              and m.mnem_friendly = 'NDWTMP16' and c.start_date  =
'9/17/98'
              and c.start_time <= '15:00:00' and c.stop_time  >=
'07:00:00') t1,
        (select start_time, stop_time
         from   events e, mnemonics m
         where  m.mnem_tag_id   = e.mnem_tag_id and   m.mnem_friendly =
'DAY'
              and   e.start_date = '9/17/98' and   e.start_time <=
'15:00:00'
              and   e.stop_time  >= '07:00:00') t2
   Where
       (t1.start_time <= t2.stop_time
        and t1.stop_time >= t2.start_time)
```

Overall features

"Overall features" of the query tool encompasses many aspects, starting from the display of the user interface, the user-friendliness of the tool, and the tool documentation. This criteria are divided into two sub-criteria; technical and non-technical.

Technical Criteria

The technical criterion consists of the following sub-criteria.

- Installation Complexity. This includes software requirements to install successfully the query tool. Ease of installation of the tool is also considered.

- Support RISQL Extensions. Since we are using Red Brick data warehouse, extending the features provided by RISQL is a plus.

- User Interface. This evaluates ease of use for the end-users in using the software to query the data warehouse. The interface should be simple and easy to use, without losing power to performing complex queries to the CCSD

- Support Aggregation. The different types of aggregation supported by the tool, such as SUM, AVG, COUNT, RANK, etc., are evaluated.

- Graphical Analysis. Software that can perform graphical-oriented analysis and reporting, in addition to the traditional reporting, is preferable.

- Query Complexity. This refers to the different types of queries, such as ad-hoc query, drill-down analysis, slice-and-dice, and drill-across, the query can support.

- Integration with Other Products. Certain users may need the ability to run software that they are familiar with on top of the query tool. For example, an end-user may want to view the output generated by the query tool using a spreadsheet program, such as Microsoft Excel. More importantly, the query tool must be able to smoothly interact with Red Brick data warehouse since the HST data are stored in a Red Brick data warehouse.

- Web Integration. With the availability of the Internet, users may want to perform query or view results of a query to the CCS Data Warehouse on the Web. The Query results generated by the tool need to be posted on the Web.

- Software administration and maintenance. This refers to the level of difficulty in maintaining the software. Prior to end-users' using the software, the different schema design, i.e. star, snowflake and fact-constellation, may need to be prepared and maintained. Query tool may provide easy-to-use and graphical-oriented software administration. Some tools may also provide the capability of administering the software and the design remotely.

- Documentation. The query tool must provide good documentation.

- Sophistication of User Required. The tool needs to be user-friendly and does not require high level of expertise of the end-users.

- Error Handling. Upon the occurrence of error, such as user, administrator or system error, how the software responds to the error is evaluated. Does the system close when there is certain error, or does the system provide notice of error to the users and have the users continue using the tool after the occurrence of errors?

Non-Technical Criteria

The non-technical criterion includes the following.

- Product Classification and Interface with Each Others

- Experience in Various Industries
- Licensing Costs
- Technical Support
- Consulting
- Online Support
- Partnerships
- Demonstration of the Software

Candidate Software

In evaluating commercial software tools, we have selected for evaluation the following major software.

- If Synchrony
- Hummingbird
- Cognos
- Brio
- Oracle Discoverer

Recommendation

In this study, four different software packages have been fully tested and evaluated based on the need of NASA HST CCS end-users. Another software package Oracle Discoverer was briefly evaluated but was discarded due to insufficient support to the Red Brick Data warehouse.

The first part of the evaluation focuses on the software support for the different relational warehouse schema designs, including the star, snowflake and fact-constellation schema. Given six different HST-related queries as sample queries, the second part focuses on the software capabilities in supporting such queries. This part of the evaluation criteria is specific to HST CCS end-users' need. The last part of the evaluation criteria consists of two different sub-criteria, the technical and non-technical features. Technical features include installation complexity, support for RISQL extensions, user interface, support for complex query, graphical analysis and non-technical features include licensing costs, use within various industries and company's partners.

Even though all of the sample queries can be designed and executed by most of the software packages, such as Brio, Hummingbird and Cognos, some of the more complex queries cannot be executed efficiently. Such complex queries can only be designed and executed through the software packages by storing results of sub-queries into new tables and later executing higher-level queries against such tables. However such an approach cannot be considered as a qualified approach to answering complex queries, where new tables have to be created prior to designing and executing the queries.

An alternative approach to answering complex queries is also considered where the HST CCS data warehouse design is modified so that it can assist the software in better designing and executing complex queries. This approach requires major changes to HST CCS data warehouse and may have to be undergone once the software package has been selected. Because of the nature of ad-hoc queries, some future complex queries may not be executed efficiently.

In addition to evaluating the different software packages, we are constantly in communication and discussion with the software vendors. Based on some of our discussion, we believe that some of the limitations mentioned in the detailed report will be resolved in future versions. NASA HST CCS could take the initiative to resolve the limitations with the selected vendor. This would enable NASA HST CCS to obtain better software packages that would enable them to design and execute complex queries effectively and efficiently. In addition, this could potentially help the vendor in developing a better software package.

In the rest of the section, we will present recommendation for each of the software packages.

If-Synchrony. If-Synchrony supports primarily the star-schema design. Even though there is an alternate approach, as discussed in the detailed report, to supporting snowflake and fact-constellation schema, they cannot be considered feasible. It is difficult to achieve some simple queries using Synchrony. The main reasons are its inability to support standard operators, such as "<=" and ">=", and its difficulty in formatting data, and its difficulty in setting conditions on the attributes of the fact table. Moreover, it does not support RISQL extension. In comparison with other software, the learning curve required for this software is longer. Even though the software is relatively new, it has some interesting features, such as slowly changing dimension. Its user interface, though easy, cannot be considered as good as the others.

Hummingbird. The software has adequate support for the three schema designs, i.e. star, snowflake and fact-constellation schema. It has a good interface and requires limited learning curve. In addition, it provides a capability to map the database design within the software where tables and their relationships can be graphically represented. Simple NASA HST CCS queries, such as queries 1,2, and 3, can be designed and executed efficiently and effectively. As mentioned in the detailed report, some complex queries, such as queries 4, 5, and 6, cannot be efficiently executed. However, sequence of tasks can be scheduled and automatically executed through the use SQL formatted queries and super queries. Moreover automation controllers like Microsoft's Visual Basic and Visual C++ can also be integrated into the system. Another feature that we feel Hummingbird to be better than others is its security feature, where data can be guarded at various different levels, such as for users, groups, objects or relationships. This enables more control over data sharing. The administrator can also set the levels of privileges (e.g., editing data models, sending SQL formatted queries or saving queries) that can be allotted to different users. Split and combined data models are introduced as described in the detail report. Such modeling allows effective manipulation of data and queries. Other features include sufficient on-line documentation, where on-line help can be customized based on user requirements, the ability to distribute queries to users for their own customization, ease of system maintenance and its wide use within various industries which shows the vendor's strength in the industry. The package does not support RISQL extensions right now but these features could be incorporated into it during the purchase of the system by their development team.

Cognos. Cognos supports the three schema designs. However, its support for fact-constellation and snowflakes cannot be considered as effective as Brio or Hummingbird. Its performance on executing queries against fact-constellation and snowflake designs is not as good as that of Brio or Hummingbird. The software requires limited learning curve and has a good user interface that allows simple queries to be designed easily. Creation of catalogs, joins, folders and classes is relatively easy. This helps in properly maintaining the data. As compared to Hummingbird and Brio, the software has limited querying capability. Security is maintained mainly at table and folder level. User can also use macros to automate tasks they include the script editor and scheduler. Among its better features is its Client/Database-server balancing option which help to optimize the processing time of the system by determining where and when the processing should occur. It also has adequate documentation and support for RISQL extension. Different types of data repository, such as snapshots, thumbnail and hotfiles, are provided. As in the case with Hummingbird, Cognos is also widely used in various industries indicating the vendor's strength.

Brio. Brio has sufficient support for the three schema designs. The software has a good interface and requires limited learning curve. It has better integration from querying to reporting to setting up charts and tables. Brio has sufficient system functions including support for RISQL extensions in comparison to other software being tested and provides end users with the ability to accomplish complex calculations. It is difficult to set interactive queries using Brio moreover the software has very limited security features in comparison to Hummingbird or Cognos. As in the case with Hummingbird and Cognos, Brio is also widely used in various industries indicating the vendor's strength.

All software packages that we have evaluated are effective in some aspects while are not as effective in others. Though none of the packages being evaluated could be recommended for highly complex querying, they are ideal for queries with simple-to-average complexity. Based on the evaluation criteria stated in the detailed report, if we were to choose one among the four, we would prefer Hummingbird. Therefore, we recommend Hummingbird software for NASA HST CCS application.

EOSCUBE: A Constraint Database System for High-Level Specification and Efficient Generation of EOSDIS Products

Alexander Brodsky
George Mason University
Department of Information and Software Engineering (ISE)
(brodsky@gmu.edu)

Summary

The EOSCUBE constraint database system is designed to be a software productivity tool for high-level specification and efficient generation of EOSDIS and other scientific products. These products are typically derived from large volumes of multidimensional data which are collected via a range of scientific instruments.

Main Objectives

- To demonstrate that EOSCUBE can provide considerable savings in development time of EOSDIS and other scientific products
- To demonstrate that product generation by EOSCUBE from real data sets is feasible.

Ultimate Goals

Productivity gain: EOSCUBE will allow Earth scientists to compactly specify data products concentrating on their scientific domains, while being relieved from a considerable programming effort.

Interleaved and Optimized Production: EOSCUBE will provide interleaved pipelined evaluation of a series of inter-related products, automatically optimizing data-flow control, buffer management, and materialization supporting clustering and indexing.

Platform Independence: EOSCUBE will support hardware/software platform independence, so that platforms' change would only require changing a small number of interface methods, while leaving products generation software unchanged. It is planned that EOSCUBE will support a mix of underlying object managers, databases, mass storage systems, or just file systems in a very flexible way.

Easy Integration: EOSCUBE is used from within a C++ program and allows to use existing C/C++ code, without the need to translate data types and formats.

Accomplishments

- Development of the EOSCUBE proof-of-concept prototype based on the CCUBE constraint object-oriented database system
- Specifying in EOSCUBE a range of scientific products, and actually generating a number of them using real input data sets.
- Preparing reports, within this final report, on:
  - Feasibility and productivity study, which contains EOSCUBE specification of a number of scientific products, and test cases run on real data sets
  - Specification of EOSCUBE features and language
  - Architecture and implementation of the EOSCUBE prototype
  - Work in progress on optimizing multi-product generation workflow
  - Recommended course of action

Main Conclusions

- EOSCUBE has the potential for significant productivity gain in specification and generation of EOSDIS and other scientific products
- Generation of scientific products from real data sets is feasible using the EOSCUBE prototype
- An industrial-strength EOSCUBE implementation will be necessary for deployment and massive use of the system.
- The EOSCUBE language should allow incremental extensions, which are unavoidable in diverse scientific domains
- The overall evaluation model should also support data-flow processing (i.e., pipeline evaluation), in addition to query processing.
- The main aspects of global optimization should deal with interleaved pipelined evaluation of series of inter-related products, and concentrate on optimizing throughput via data flow control, buffer management, and materialization supporting clustering and indexing.

Future Action Paths for EOSCUBE

We elaborate on recommended activities in Section VI. Below is a summary of main paths of action that will have to be carefully discussed and planned with EOSDIS.

Research Path, including local and global optimization, spatio-temporal indexing and clustering, and GIS constraint algebras

Industrial-strength implementation path, including high-performance EOSCUBE kernel, pipeline evaluation model, ODBC and platforms support, and GIS integration.

Collaborative work with Earth scientists on a specific set of new products, and continued customization of EOSCUBE for them. This will also used as a leverage for later massive deployment of EOSCUBE.

Deployment of EOSCUBE to Centers and Technical Support

Reference

Brodsky, A. and Segal, V. (1999). *EOSCUBE: A Constraint Database System for High-Level Specification and Efficient Generation of EOSDIS Products.*

Mass Storage Performance Information System

Bin Chen
Northwestern University
Department of Electrical and Computer Engineering
(bchen@ece.nwu.edu)

1. Introduction

Detailed logs capture the activity of Mass Data Storage and Delivery System (MDSDS), e.g. data being transferred, the location of data on the server, the transfer speeds, the users who access the system, etc. These logs are essential in analyzing the system performance and security. However, the lack of structures, missing values, inaccessibility to querying, and inconsistent log data make the log records difficult to use, especially it is very time consuming on creating reports about system usage. Therefore, an urgent need in system maintenance exists for cleaned and organized log records.

In the past year, a Mass Storage Performance Information System (MSPIS) has been created to facilitate the organization of log data created by the MDSDS at NASA Science Computing Branch. The MSPIS is also designed to aid the discovery of knowledge from cleaned log data, for example, the average mount time of each tape drive, the comparison of tape read/write speeds, the time of data when most requests occur. Such information is of great help in improving system performance.

The project of setting up a MSPIS for NASA CESDIS can be divided into two phases. In the first phase, which spanned from 1997 to 1998, a front-end database and a set of data extraction and manipulation tools were designed by Lisa Singh of Northwestern University. The data extraction and manipulation tools take data from original log files, clean the raw data, and store them into a front-end database. In the second phase, we try to move the historical data into a data warehouse and discover knowledge from the historical records. From January 1999 to June 1999, a data warehouse was designed and built by the author. Most experimental data have been successfully moved from the front-end database into the warehouse. By using the historical data, the user access patterns can be analyzed, as well as their changing trends. Moreover, the relationships between data movement and date/time can also be revealed.

This report is organized as follows. In section 2, the overall system design is introduced. Work that has been done from January 1999 to June 1999 will also be summarized. Section 3 discusses future work. Finally, section 4 concludes the report.

2. System Architecture

2.1 Mass Data Storage and Delivery System (MDSDS)

The Science Computing Branch at NASA Goddard manages the world's most active openly accessible storage system, supporting over 800 local and remote users. The MDSDS runs Uni-Tree software. The software manages over 2 million files with an average size of 19 megabytes. The user data totals almost 50 terabytes and grows approximately 1 terabytes per month.

The UniTree processes write messages to various log files. The log files associated with the MDSDS have been saved for more than 5 years. Since these log files attempt to reflect all changes and accesses made to the system, they grow rapidly. There are four different types of logs generated by the UniTree software: ftp, mnt, pdm, and utm. The ftp log files contain detailed information on all ftp sessions, including user id, transferred file names, sizes of transferred files, etc. The pdm logs maintain data about tape mounts, e.g. mount duration and tape drive access time. The mnt logs contain details about all mount/dismount operations and all searches for available tape drives. Finally, the utm logs record all the UniTree demon output messages, including process details, process duration, transfer rates, number of bytes transferred, etc.

2.2 Mass Storage Performance Information System (MSPIS)

The Mass Storage Performance Information System has been constructed to facilitate knowledge discovery and information querying from the MDSDS log data. It is an essential component of the decision support system to analyze the system performance. The MSPIS contains two parts: The first part includes a front-end database, as well as the tools to extract and clean raw log data. The second part consists of a data warehouse and the data analysis tools. Figure 1 shows the architecture of such a system.

The data extraction and manipulation tools extract log records from log files. Records extracted can be ill-structured. The data manipulation tools clean all the records: missing values are identified, inconsistent fields are corrected, and related records from heterogeneous systems are linked. A front-end database is then used to store the processed log information. The use of such a front-end database provides us an intermediate storage. While log files are being continuously updated, the data cleaning and organization work can be done on-line by using such an intermediate database and some temporary files. Since the data extraction and manipulation tools were implemented between 1997 and 1998, we will not detail the design in this annual report. For detailed information of the front-end database design, please refer to Lisa Singh's annual report of 1998.

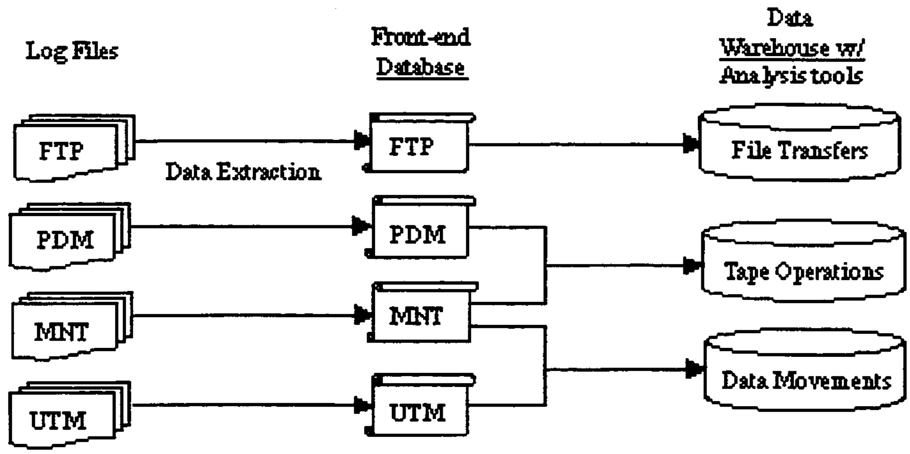


Figure 1: Mass Storage Performance Information System (MSPIS)

Compared to the front-end database, which is frequently updated to include newly extracted log information, the data warehouse is a relatively passive and static database. Data stored in the front-end database are periodically moved into the warehouse. Hence the data warehouse is a historical database. We plan to move all the five-year log data into the warehouse.

A constellation structure is used to construct the warehouse, as shown in Figure 2. The constellation structure consists of three star structures, *File Transfers*, *Data Movement*s, and *Tape Operations*, linked by a *Time* table, which records all dates, days of week, hours, minutes and events. The *Time* table is shared by all star structures. Each star structure is in fact a sub-warehouse and stores data about one specific operation. For example, *File Transfers* division holds all the historical data on user file transfers, *Data Movements* division records all the file movement activities in the MDSDS, and *Tape Operations* tracks all the tape mounts, dismounts and searches for available tape drives.

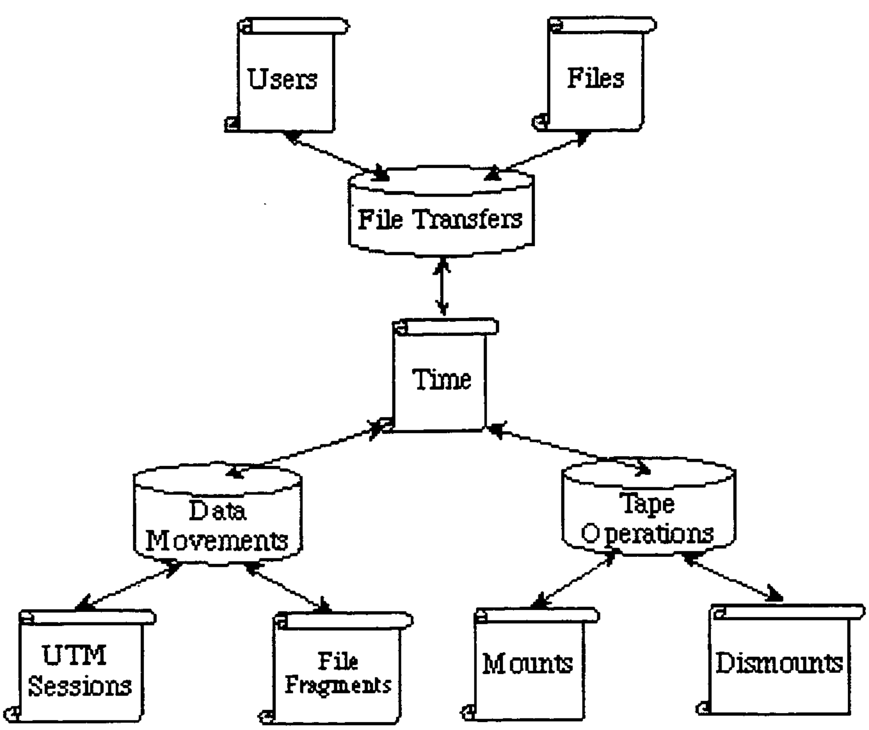


Figure 2: Data Warehouse Design

A star structure can be broken down into a fact table and a number of dimension tables, e.g. *Users, UTM Sessions*, and etc. The fact table contains detailed records, especially all the measurable fields that may appear in user queries, e.g. transfer speed, duration, bytes moved, etc. Each dimension table consists of some closely related fields. The fields in each table were carefully selected to reflect as many user operations and data movements as possible. To quickly answer queries, each table in the data warehouse was assigned a dimension key, or surrogated key. The reasons to use surrogated keys instead of natural keys are that keys are no longer related to data types of natural keys and the same type of keys can be used in any table regardless of the contents of the table. The independence of keys from data types enables us to always use integers as the surrogated keys. Because the comparisons of integers are normally much faster than those of strings and floating numbers, to join tables by integral keys can result in a much better performance and compact representation. Moreover, if surrogated keys are used in every table, we can join tables without concerns that the data types and formats of keys are inconsistent among records from different sources since log data may be extracted from heterogeneous systems.

To achieve better performance, a large number of indices have been built. For ordinary databases, although indices can speed up the query process, there are also some disadvantages, e.g. the storage overhead for indices and the slow down during updates. However, because the data warehouse is designed to handle a large amount of data, the storage spent on indices becomes trivial compared to the size of data in the warehouse. As to the speed of updates, normally, the more indices, the worse performance on updates. While since the warehouse is a historical record, instead of an ordinary database or an OLTP system, updates are much less frequent and more non-time-critical queries are expected instead. This feature makes the improvement in query answering offset the disadvantages caused by indices. I have already loaded some log data that is equivalent to one quarter of log records. From some experiments on joins, we verified that the introduction of such indices did improve querying speed significantly. Meanwhile, the performance of warehouse updates was still tolerable.

To make the warehouse more accessible to general users, a friendly user graphical interface is being built. By using this interface, end users can submit queries by simply filling out a query form. Knowledge of SQL is no longer a requirement for the users.

The warehouse not only can store a large amount of historical data, but also is able to offer decision support. To convert a warehouse into an intelligent decision support system, two areas have been targeted. First, the warehouse should be able to answer a number of frequently asked queries. Secondly, tools for knowledge discovery from large amount of data should be available to decision-makers.

The frequently asked queries were identified and supplied by the MDSDS administrators. I embedded all the corresponding SQL blocks to the warehouse. If a user wants to find the answer to a query, the warehouse can run the corresponding SQL blocks. Related tables will be searched and answers will be returned to the user. In this case, the indices mentioned before are especially useful. Most queries can be answered by using indices.

The knowledge discovery will be implemented by a set of data mining tools. However, currently we are still focusing on the construction and testing of the warehouse. Those tools will be discussed in next section.

3. Future Work

The next step for the project is to design a set of data mining tools. Usually there is rich hidden knowledge in warehouses. As previously stated, it would be of great help for system administration if user access patterns, the relationships between the number of accesses and date/time, and etc. can be identified from the warehouse. Although much research has been done on the knowledge discovery from large databases, little has been accomplished in the area of data warehouses. The difficulties on mining from warehouses stem from the fact that warehouses are normally much larger than traditional databases and for performance consideration, most of them are not normalized. To make things worse, traditional SQL does not support knowledge discovery from data warehouses and databases.

The warehouse we designed reflects the above features. For example, redundant fields are included in some dimensional tables to facilitate the query process. Some aggregate values are also pre-computed and stored in the warehouse. The non-normalized data make the knowledge discovery procedure more time-consuming. However, there are also some advantages in mining from data warehouses: The pre-computed aggregate values may help us save time on mining generalized association rules, which are frequent patterns over hierarchical data.

To decide which tools we should design, the feedback from system administrators and the data warehouse content should be carefully researched. The possibly useful mining tools for NASA MDSDS data include the classification rule mining module to discover user access patterns and to classify user data and the sequential and association rule modules to find out which files are usually accessed together.

4. Conclusion

The design of Mass Storage Performance Information System for NASA has been advanced successfully for more than one year. During the past six months, a warehouse has been set up to host the historical log records. Some experimental data have been loaded and many experimental queries have been implemented and tested. A graphical user interface is being built to make the warehouse more accessible to end users. In the near future, some data mining modules will be added to the system.

Konstantinos Kalpakis
University of Maryland Baltimore County
Department of Computer Science and Electrical Engineering
(kalpakis@cs.umbc.edu)

Worked described below by Dr. Kalpakis is supported by a cooperative agreement from NASA to USRA, managed by CESDIS for USRA.

My primary focus during this period was development work for the Environmental Legal Information System. I undertook a number of developed activities related to the ELIS project. We developed a prototype system architecture. The main line of the approach was to utilize the services provided by traditional relational database management systems, and the Arc/Info Geographic Information System, together with Web and Extensible Markup Language technologies. I developed a database for the legal texts and loaded all the Global Legal Information Network (GLIN) documents, as well as additional documents from US Laws. A major activity was the design of an XML document type definition (DTD) for legal texts incorporates and links legal texts with geographic and environmental information. The developed DTD subsumes the GLIN schema. To demonstrate that, I wrote a set of scripts that convert GLIN data into the format required by this

DTD. Further, I developed a prototype Java XML editor for use with the DTD. Subsequently, I customized another XML editor from IBM for that purpose that seemed more appropriate for this task by my colleagues. Additional effort is required. The major activity was to develop a set of Java servlets for enabling access to the ELIS prototype through the Web. This effort lead to a number of benefits that were not realized in earlier GLIN prototypes. For example, the new system has significantly smaller response times, it reduces the load imposed on the Web servers, it eliminates some significant security issues with respect to the database, and simplifies the installation tasks at clients. Further, it is much easier to extend and maintain.

Subsequently, I looked into further simplifying the development of such servlets; currently, I am examining Java Active Pages technologies and their use with servlets as the main technologies for building the ELIS system. At the same time, I have a number of activities related to XML and databases. Currently, I am looking into methods for ingesting XML objects into Object-Relational databases (e.g. Oracle 8 in our case). I am also developing Map Services using ESRI's MapObjects technologies to enable the presentation of GIS data within ELIS. This effort is ongoing, a small prototype has been developed, and I am in process of integrating it with the rest of the system. Further, I designed a new algorithm for optimally placing copies of files on the nodes of a network that take into account not only read and write costs, but also storage costs and capacity/load constraints at the nodes. A paper has been submitted for publication to IEEE Transactions and is currently under review. Finally, I participated in two WP-ESIP federation meetings were we presented the status of ELIS, and participated in the Interoperability Working Group of the federation.

DVNS Science Applications
(in collaboration with Task 75- Jeanne Behnke and Joel Sachs)

We have been working on an experimental evaluation of Informix Datablades for very large spatial databases. In particular, we are working with the MONET catalog of stars (which has about 500,000,000 objects), to investigate the suitability of existing datablades (e.g. GeoDetic for Earth Science Data and Shapes2 for traditional geometric data) for on-line spatial queries, as well as decision-support queries. In the context of this effort, some extensions to those datablades have been made. One of the objectives of this effort is to examine the efficacy and efficiency of commercial Object-Relational Database Management Systems for science data (note that the MONET catalog is accessible through specialized programs) and analyze the benefits and trade-offs of such approaches.

Scalability Analysis of ECS's Data Server

Daniel A. Menascé
George Mason University
Department of Computer Science
(menasce@cs.gmu.edu)

During this year we continued and expanded the goals of the study carried out last year with respect to analyzing the performance and scalability characteristics of EOS Core System's Data Server.

During the year the following tasks were accomplished:

1. Prepared a complete written report of the results obtained until December 31, 1999.

2. Prepared a set of programs to automate the use of the ECS Scalability Analyzer in studying several scenarios.

3. Analysis of Compression Techniques: We used the Scalability Analyzer to verify different scenarios in which compression could be used in ECS's Data Server. Experiments and measurements of four compression algorithms on several types and size of files were carried out by Pen-Shu Yeh. Using these data, we built several scenarios for the use of compression including distribution in compressed form (DC) and distribution in uncompressed form (UD). The several studies showed that the use of DC along with the compression algorithm called sz, developed by Yeh, provided the best performance and allowed reprocessing to take place with the current configuration.

4. Bottleneck Analysis: A preliminary analysis of bottleneck removal was done and it was determined that the distribution server would be the bottleneck in the UD scenarios. A four times faster distribution server would solve the problem. Further bottleneck analyses will be carried out taking into account the new configuration for GSFC's DAAC.

5. Use of Optical Tapes: A preliminary analysis of the use of optical tapes and fewer tape drives with the faster tapes was started. Preliminary results show that the faster tapes allow for a smaller number of tape drives to be used while maintaining equivalent performance.

Publications

Pentakalos, O, Menascé, D. A. and Yesha, Y. Analytic Modeling of Distributed Hierarchical Mass Storage Systems with Network-Attached Storage Devices, to appear in the *IEEE Transactions on Parallel and Distributed Systems.*

Menascé, D. A., Almeida, V., Fonseca, R. and Mendes, M. A. (1999). Resource Management Policies for E-commerce Servers. *Second Workshop on Internet Server Performance, in conjunction with ACM SIGMETRICS 99/FCRC*, Atlanta, GA.

Harbitter, Alan and Menascé, Daniel (1998). Performance Issues in Large Distributed System Security. *Proc. of the Computer Measurement Group Conference*, Anaheim, CA.

Foxwell, Harry and Menascé, D. A. (1998). Prefetching Results of Web Searches. *Proc. of the Computer Measurement Group Conference*, Anaheim, CA.

Menascé, D. A. and Kakarla, Nagesh (1998). Performance and Availability Modeling of Replicated Database Servers. *Proc. of the Computer Measurement Group Conference*, Anaheim, CA.

Dodge, R. and Menascé, D. A. (1998). Prefetching Inlines to Improve Web Server Latency. *Proc. of the Computer Measurement Group Conference*, Anaheim, CA.

Menascé, D. (1998). Web Performance Modeling Issues. *First NASA Information Systems Performance Modeling Workshop*, NASA Ames Research Center, Moffett Field, CA.

Menascé, D. A. (1998). Educational Challenges and Opportunities in the Web Era. *Proc. of the Workshop on CS Education, 1998 Annual Congress of the Brazilian Computer Society*, Belo Horizonte, Brazil, (invited paper).

Menascé, D. (1998). Web Performance: the Most Obvious Challenge to Electronic Commerce. *Proc. of the 1998 Annual Congress of the Brazilian Computer Society*, Belo Horizonte, Brazil, (invited paper).

Menascé, D. A. and Gomaa, H. (1998). On a Language Based Method for Software Performance Engineering of Client/Server Systems. *First International Workshop on Software Performance Engineering*, Santa Fe.

Menascé, D. A. and Almeida, V. A. F. (1998). *Capacity Planning for Web Performance: metrics, models, and methods*, Prentice Hall: Upper Saddle River, New Jersey.

Advanced Geo-stationary Study

Tarek El-Ghazawi
George Mason University
(Tarek@gmu.edu)

Image Resampling for AGI

Image resampling is the process of extrapolating data values to a new grid. Thus, resampling refers to calculating pixel values for the rectified grid from the original grid. In the AGI, the data pixels obtained from horizontal and vertical movements of the scanning mirror, which corresponds to the inner and outer variations of the scanning angle, must be resampled to a rectangular grid producing digital images ready for manipulation. Many resampling techniques have been considered for the AGI study. Many have been tested to delineate their ability to preserve key characteristics of the original image, such as radiometry and geometry. Due to the overlapping between swath data, the resampling techniques considered will only require data from the same swath for producing the final pixel data. Section 1 overviews the most promising resampling techniques. Section 2 overviews some of the results obtained to quantitatively assess the resampling methods. Section 3 provides some future directions for the next phase of this research. Detailed results can be found in the appendix.

1. Resampling Methods

A number of candidate resampling techniques were considered and compared. These include: Nearest Neighbor, Bilinear interpolation, Cubic Convolution, Hanning Windowed Sinc, and Optimal Deconvolution. All of these methods were examined conceptually. Full implementations for the first three methods were conducted and their relevant comparative results will be provided in the next section. Due to the time limits of our study, no sufficient time and/or details were available to implement and test the last two techniques.

1.1 Nearest Neighbor

Nearest neighbor approach uses the value of the closest input pixel for the output pixel. Advantages include the fact that method is very fast, and output values are the original input values. The last point is particularly important when pixel values are needed for classification, such as determining vegetation types. The major disadvantages are the fact that data values are lost, which also leads to the production of a choppy stair-stepped effect in the image. Computational requirements of this method are in the neighborhood of 4 FLOP/Pixel.

1.2 Bilinear Interpolation

This method uses a weighted average of the nearest four pixels to produce the output pixel. The major advantage is reducing the stair-step effect caused by the nearest neighbor approach. However, this method alters the input pixel values and its smoothing effect reduces the contrast and the high-frequency content of the image. The method could potentially create dark cells around the perimeter of the output file. In addition, the bilinear method is more computationally expensive than nearest neighbor. Computational requirements of this method are in the neighborhood of 40 FLOP/Pixel.

1.3 Cubic Convolution

Cubic convolution uses a 2D polynomial approximated over sixteen nearest pixels to produce the output pixel. This eliminates the stair-step effect caused by the nearest neighbor approach. Contrast is still reduced than the nearest neighbor, but better than bilinear. It is accurate, however, for low spatial frequency. This algorithm is almost an order of magnitude more computationally expensive than the bilinear method. Computational requirements of this method are about 200 FLOP/Pixel.

1.4 Hanning Windowed Sinc

This algorithm depends on optimally windowed truncation of IIR Sinc interpolator. It distributed the error uniformly over the frequency space. It is, however, a much slower algorithm than those listed above. Computational requirements of this method are estimated at over 414 or 1150 FLOP/Pixel for 3x3 and 5x5 window sizes, respectively.

1.5 Optimal Deconvolution

This method uses a Weiner filter to determine appropriate resampling kernels. These kernels are to remove, upon deconvolution, the effects of the imaging instrument (blur and phase distortion). The method, unlike these previously discussed, requires a prior knowledge of the instrument characteristics.

2. Comparative Results

	Radiometry	**Geometry**	**Blur**	**Computational Complexity**
Nearest Neighbor	Best	Very bad	Best	Least Computationally Demanding
Bi-linear	Worst	Good	Bad	2^{nd} Best
Cubic Convolution	Good	Better	Good	3^{rd} Best
Hanning	Good	N/A	Very Good	4^{th} Best
Optimal Deconvolution	Very Good	Best	Very Good	Most Computationally Demanding

Table 1: Subjective Comparison of Resampling Methods

TM	Image	Histogram
ORIG-NEAREST	18.83	579.58
ORIG-BILIN	24.73	1521.94
ORIG-CUBIC	17.3	626

Table 2: RMS Error results

Table 1 presents a subjective comparison of the five examined resampling techniques. Some quantitative insight is provided in Table 2. This table summarizes the results from correcting a thematic mapper image, which is rotated by 18 degrees, using the three listed resampling techniques. The root mean squared error between the original image and the corrected version is given in the second column. The third column gives the rms value between the histogram of the corrected and the histogram of the original image. The rms of the histograms focuses on changes in radiometry, while the rms for the images takes into account other factors, such as geometry. The results confirm that while the nearest neighbor approach is best for radiometry, the cubic convolution is generally the better one.

3. Recommendations and Future Directions

There was not sufficient details and time during this study to produce empirical results for the Hanning and Deconvolution techniques. The relative performance of these two methods is reported in Table 1 only based on the general conceptual belief. These two methods seem to be promising and future studies should consider benchmarking them against the other three popular methods of Table 2. The real potential for these methods can be fully understood if an experimental study including all five methods were to be conducted.

For the three popular methods of Table 2, cubic convolution seems to be best. However, its major problem is the smoothing effect which changes the image radiometry. It seems from our study, that one can construct a new dual track method which could use nearest neighbor results to improve the radiometry of the cubic convolution, resulting in a fast and more accurate method.

Appendix

Additional results are given in this appendix. These and previous results are based on two sets of images. The first was a digitized photo of a girl, called the girl image, and was rotated by 18 degrees. The second, was a thematic mapper image rotated by four degrees. The resampling methods were used to correct the rotated images and map them back to the original grid. The root mean squared error was used, as explained before, to compare the corrected images and the original (correct one).

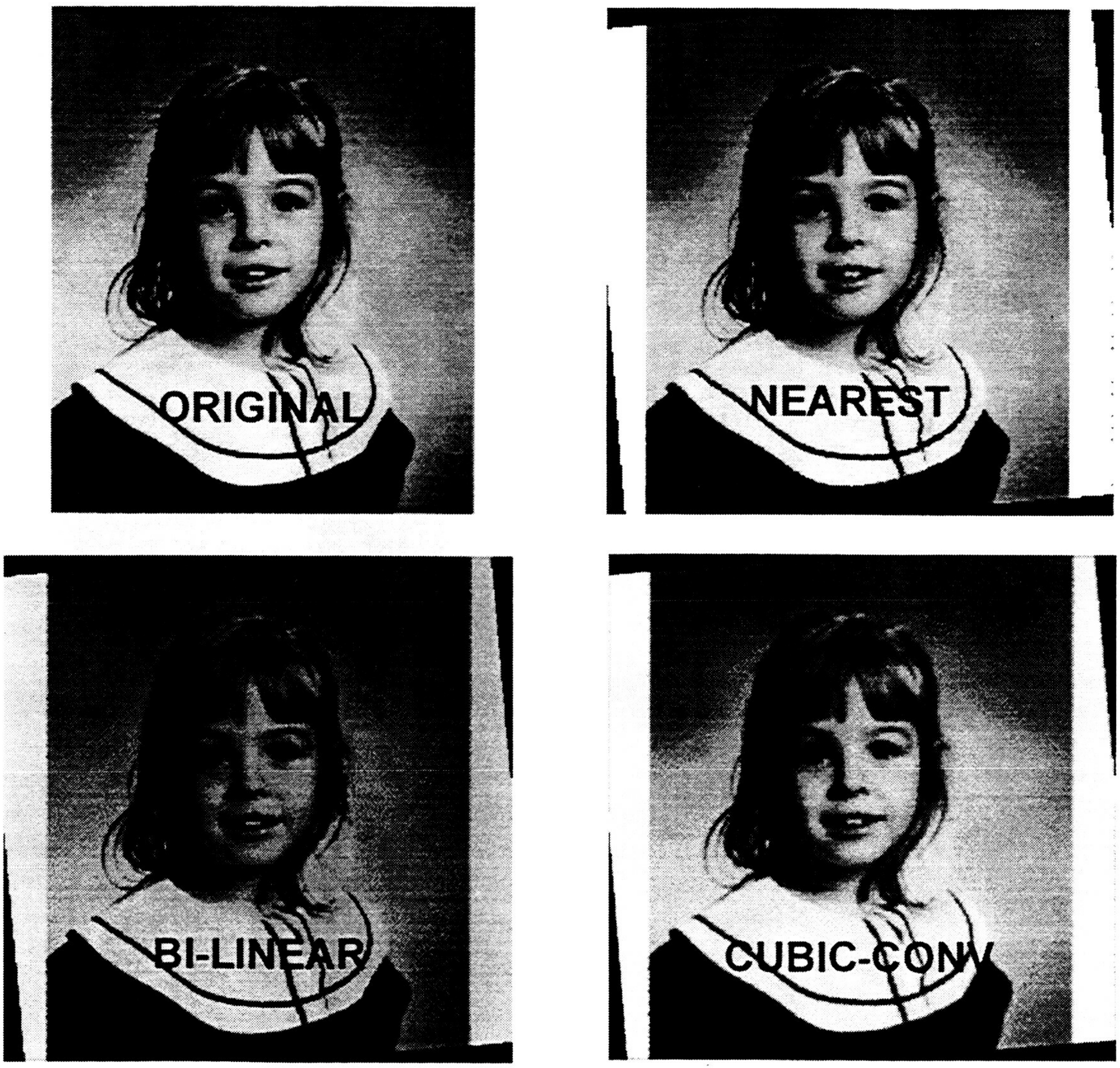


Figure 1: GIRL resampled images using Nearest Neighbor Bilinear Interpolation, and Cubic Convolution

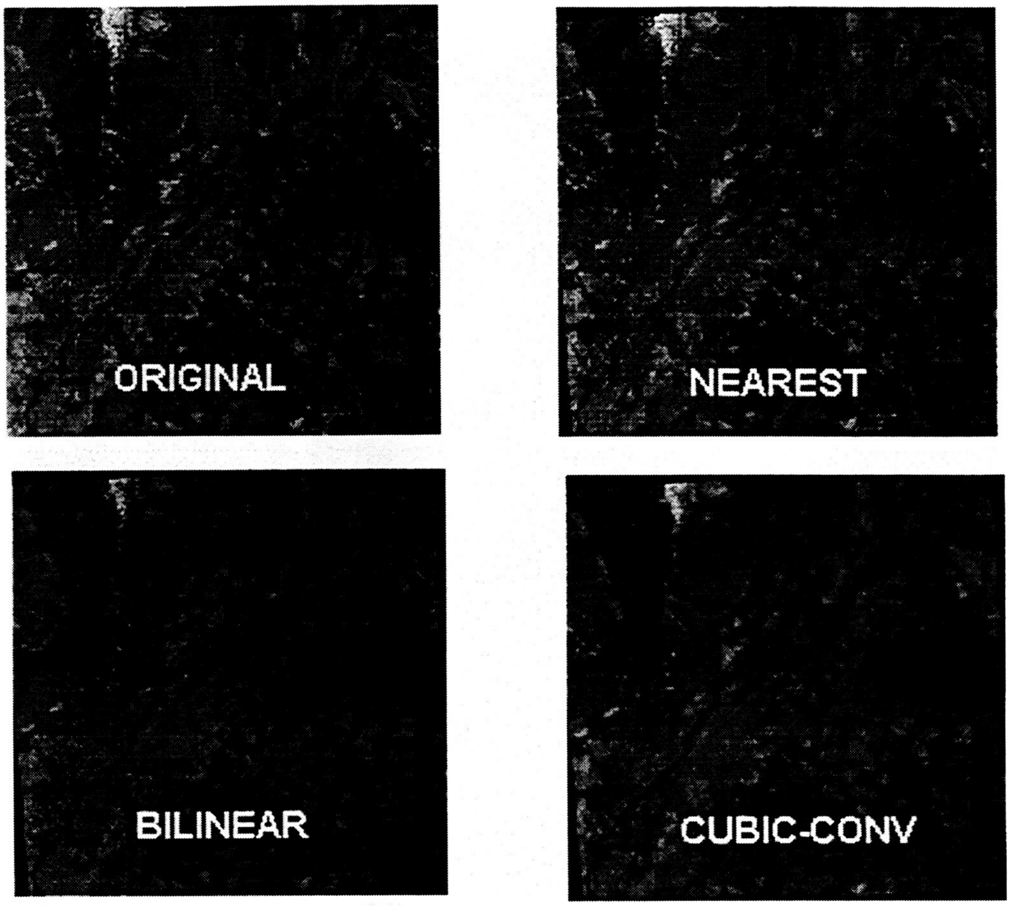
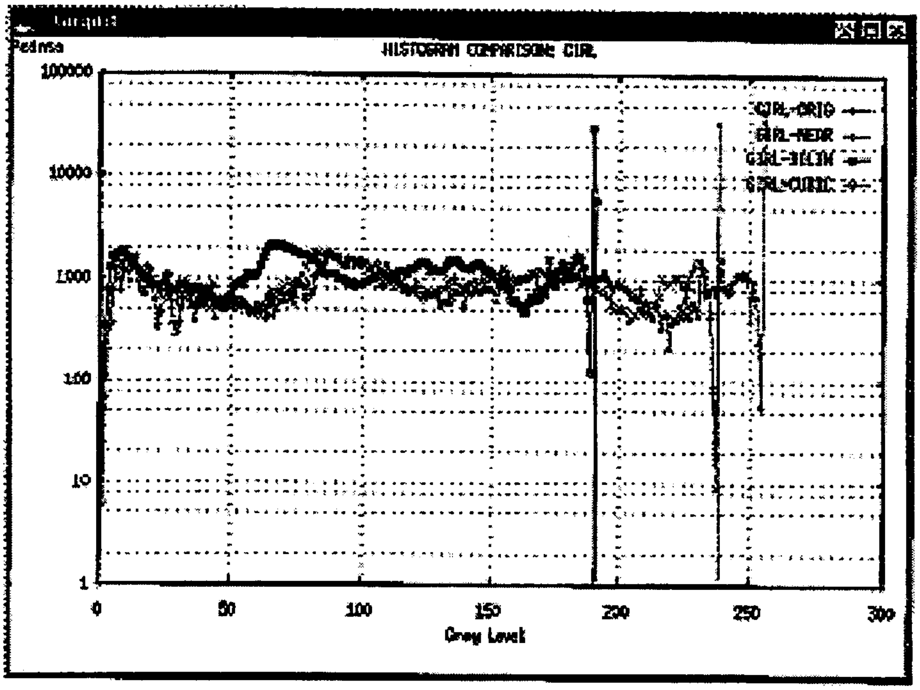


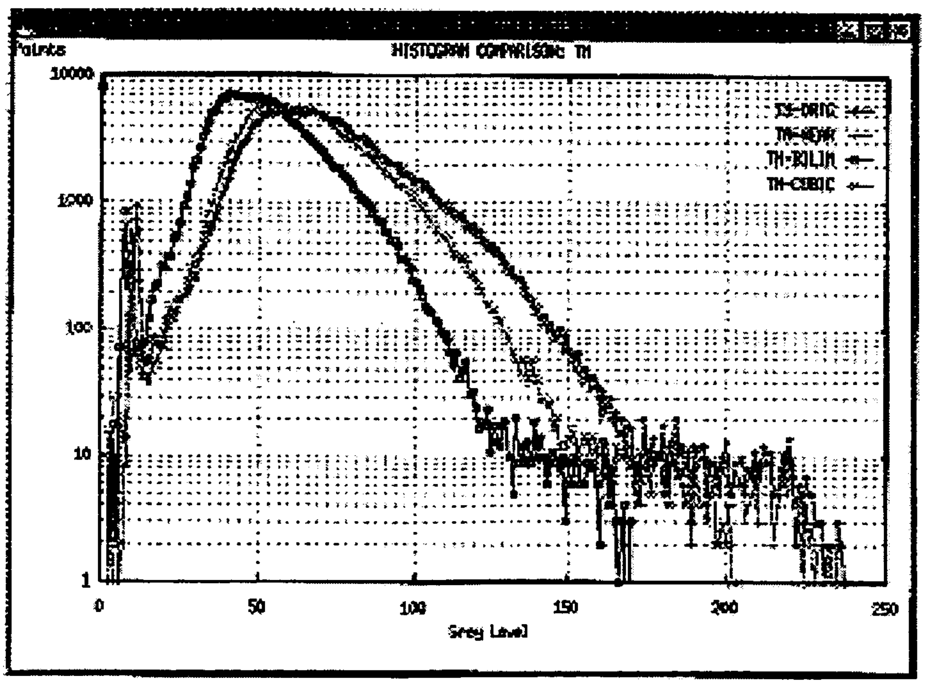
Figure 2: TM resampled images using Nearest Neighbor, Bilinear Interpolation, and Cubic Convolution

GIRLS	**Image**	**Hist**
ORIG-NEAREST	46.04	**835.2**
ORIG-BILIN	59.94	3441
ORIG-CUBIC	**44.92**	3519
TM		
ORIG-NEAREST	18.83	**579.6**
ORIG-BILIN	24.73	1522
ORIG-CUBIC	**17.3**	626

Table 3: Root Mean Squared Error for GIRLS and TM images

Histogram Comparison





An Evaluation of Automatic Image Registration Methods

– Final Report –
Jacqueline Le Moigne
(lemoigne@backserv.gsfc.nasa.gov)

In studying how our global environment is changing, research in Earth Science involves the comparison, fusion and integration of multiple types of remotely sensed data at various temporal, radiometric and spatial resolutions. Results of this integration will be utilized for global change analysis, as well as for the validation of new instruments or of new data analysis. In order to help such activity, my research work has focused on image geo-registration as well as feature (or content) extraction.

Image Geo-Registration

Digital image registration is very important in many applications of image processing, such as medical imagery, robotics, visual inspection, and remotely sensed data processing. For all of these applications, image registration is defined as the process which determines the most accurate match between two or more images acquired at the same or at different times by different or identical sensors. Registration provides the "relative" orientation of two images (or one image and other sources, e.g., a map), with respect to each other, from which the absolute orientation into an absolute reference system can be derived. My work in the registration domain has focused on surveying all the different techniques used for image registration, on developing a new method based on wavelet transforms and on evaluating different methods by the means of a toolbox.

In the following, image registration will be defined with one set of data taken as the *reference data*, and all other data, called *input data*, matched relative to the reference data. According to previous surveys on registration, data registration can be viewed as the combination of four components:

1. *a feature space,i.e.* the set of characteristics used to perform the matching and which are extracted from reference and input data,

2. *a search space, i.e.* the class of potential transformations that establish the correspondence between input data and reference data,

3. *a search strategy,* which is used to choose which transformations have to be computed and evaluated,

4. *a similarity metric,* which evaluates the match between input data and transformed reference data for a given transformation chosen in the search space.

The transformation which gives the best match according to the similarity measure is also called the *deformation model.* According to some a priori knowledge of the data, different search spaces may be chosen. Some common transformations that are often used are *rigid transformations* (composed of a scaling, a translation and a rotation), *affine transformations* (composed of a rigid transformation, a shear and an aspect-ratio change), and polynomial transformations. The most common approach to registration is to reduce the feature space to a few outstanding characteristics of the data (for example, known geographic features), which are called *ground control points* (GCP's) or *reference points.* Then the GCP's are used to compute the coefficients of a bivariate polynomial, usually of degree 3 maximum. The similarity metric in this case is a least mean-square estimator. Most commercial systems assume some interactive choice of the GCP's, and are not well suited for the automatic processing of a large number of data. Thus, the main issues related to this approach are first to extract some GCP's quickly and automatically, and second to match these points one to one. Often the deformation model is then computed as a polynomial transformation. But polynomials require either a few very accurate control points or many inaccurate ones. In the last case, the fitting of the bivariate polynomial by the least mean-square method could be very time consuming. To solve this problem, this research work has been focusing on methods which have the potential to be fast and accurate.

Wavelet-Based Image Registration

The Wavelet Transform, similarly to the Fourier Transform, is very useful to perform signal analysis and reconstruction, and especially to analyze 2-D images. Wavelet transforms provide a time-frequency representation of a signal, which can be inverted for later reconstruction. However, the wavelet representation allows a better spatial localization as well as a better division of the time-frequency plane than a Fourier transform, or than a windowed Fourier Transform. In a wavelet representation, the original signal is filtered by the translations and the dilations of a basic function, called the "mother wavelet". For our registration study of two-dimensional remote sensing images, we will only consider discrete orthonormal basis of wavelets. This choice will allow us later to tie this algorithm with a more general data management framework, in which wavelet decomposition could serve the multi-purpose of data registration, data compression, data reconstruction, and feature extraction for further analysis.

In this work, I showed how maxima of wavelet coefficients can form the basic features for an automatic registration of multiple resolution data.

Following the registration framework described in the previous section, our algorithm utilizes the four following components:

1. *The feature space*
   According to Mallat's algorithm, an orthonormal basis of wavelets can be defined by a scaling function and its corresponding conjugate filter. In this case, the wavelet decomposition of an image is performed in a multi-resolution fashion and is similar to a quadrature mirror filters decomposition with the low-pass filter L and its mirror high-pass filter H: each image is filtered in rows and then in columns by the two filters before being decimated by 2 in each direction. Then the process is iterated by decomposing again the "compressed" subimage or low-frequency subband. We will call LL, LH, HL, and HH the four images created at each level of decomposition, where LL is the compressed image and {LH,HL,HH} are the detail subimages corresponding to high-frequency components. From previous experiments performed on images of human faces, we found out that the two images LH and HL contain the most significant features, similar to edge features. Therefore, we chose to use only those features in the registration process. After computing the histograms of these two images, we only keep the points whose intensities belong to the top n% of the histograms (n being a parameter of the program whose selection can be automatic); we call these points "maxima of the wavelet coefficients" (or "maxima"). These maxima are computed for all levels of the wavelet decomposition, for reference as well as input images.

2. *The search space*
   In a first step, we assume the transformation to be either a rigid or an affine transformation. Both types of transformations include compositions of translations and rotations; therefore, as a preliminary study, our search space is composed of 2-D rotations and translations, and will be extended later to rigid and affine transformations. We look for rotations with angles included in the interval [0,90degrees] and for translations in the interval [0, half pixel-size of reference image].

3. *The search strategy*
   Our search strategy follows the multi-resolution wavelet decomposition, starting at the last level of decomposition and going back up to the first level of decomposition, i.e. going from low resolution up to high resolution. After the maxima of the wavelet coefficients of both reference and input images have been computed for all levels of decomposition, the maxima of the reference image are successively transformed by all the transformations included in the search space. The accuracy of this search increases when going from low resolution to high resolution. At each level, the search focuses in an interval around the "best" transformation found at the previous level. See Table 3.2 for a summary of this search strategy when registering 512x512 images.

4. *The similarity metric*
   At each level of decomposition and for each of the transformations, a correlation measure is computed between transformed reference maxima and input maxima.

After developing a parallel implementation of wavelet decomposition on a Single Instruction Multiple Data (SIMD) massively parallel computer, the MasPar MP-2, this wavelet-based registration algorithm was tested successfully with data from the NOAA Advanced Very High Resolution Radiometer (AVHRR), the Landsat/Thematic Mapper (TM) as well as from the Geostationary Operational Environmental Satellite (GOES). Results are summarized in CESDIS Technical Reports 94-112, 95-146, and 96-182, as well as in the 1994, 1995, and 1996 Annual Reports.

Image Registration Toolbox and Evaluation of Image Registration Techniques

(in collaboration with W. Xia -GST, J. Tilton -Code 935, P. Chalermwat and T. El-Ghazawi -GMU, N. Netanyahu and D. Mount -UMD)

As this need for automating registration techniques is recognized, each new program involved in the development of a new instrument is independently developing another registration method. Very often, these methods are developed based on something quite similar existing for another sensor, without surveying all the possibilities. Therefore, we feel that there is a need to survey all the registration methods which may be applicable to Earth Science problems and to evaluate their performances on a large variety of existing remote sensing data as well as on simulated data of soon-to-be-flown instruments. In this work, we have: 1) developed an operational toolbox which consists of several registration techniques, and 2) provided a first quantitative intercomparison of the different methods, which will allow a user to select the desired registration technique based on this evaluation and the visualization of the registration results. Results are summarized in CESDIS-TR-98-221 and in the 1997 and 1998 Annual Reports.

Feature Extraction

As the amount of multidimensional remotely sensed data grows tremendously, Earth scientists need more efficient ways to search and analyze such data. In particular, extraction image content is emerging as one of the most powerful tools to perform data mining. Some of the most promising methods to extract image content are image segmentation, which provides a spatial description of the images into parts (objects or regions), or image classification, which provides a labeling of each pixel in the image. Segmentation can be performed in several ways, which are categorized as pixel-based, edge-based, and region-based. Each of these approaches are affected differently by various factors, and the final result may be improved by integrating several or all of these methods, thus taking advantage of their complementary nature. In the following works, I first consider an approach that integrates region growing segmentation and edge detection results by interpreting a binary tree representation, thus producing a refined region segmentation. I also proposed to perform the integration of edge and region data by a relaxation method; in the research described here, this relaxation method is refined for the purpose of integrating edge and classification information and is implemented on a massively parallel computer, the MasPar MP-1. Finally, a method integrating neural network classification with wavelet processing is investigated.

Integration of Edge Information to Image Segmentation

(in collaboration with J. Tilton -Code 935)

Image segmentation is often one of the first steps in the analysis of remotely sensed data. This work focuses on two particular types of segmentation: region-based and edge-based segmentations. Each approach is affected differently by various factors, and both types of segmentations may be improved by taking advantage of their complementary nature. Included among region-based segmentation approaches are region growing methods, which produce hierarchical segmentations of images from finer to coarser resolution. In this hierarchy, an ideal segmentation (ideal for a given application) does not always correspond to one single iteration, but may correspond to several different iterations. This, among other factors, makes it somewhat difficult to choose a stopping criterion for region growing methods. To find the ideal segmentation, we develop a stopping criterion for our Iterative Parallel Region Growing (IPRG) algorithm using additional information from edge features, and the Hausdorff distance metric. We integrate information from regions and edges at the symbol level, taking advantage of the hierarchical structure of the region segmentation results. Also, to demonstrate the feasibility of this approach in processing the massive amount of data that will be generated by future Earth remote sensing missions, such as the Earth Observing System (EOS), all the different steps of this algorithm (namely, region growing, edge detection, Hausdorff distance computation, and edge/region fusion) have been implemented on a massively parallel processor. Results are summarized in CESDIS-TR-95-146.

Integration of Edge Information to Image Classification

A large number of iterative relaxation schemes have been proposed to improve the results given by such basic processes as edge detection, region segmentation or pixel classification. The principle of these algorithms is to utilize contextual information for iteratively changing the initial labeling of the objects in a scene toward optimal labeling. In this work, only relaxation methods, for which the decisions at each point are taken in a probabilistic fashion, are considered. Such a method is utilized to integrate knowledge from edge detection and pixel classification. Results are summarized in the 1993 CESDIS Annual Report and in CESDIS-TR-93-95.

Integration of Wavelet Information to Image Classification

(in collaboration with N. Netanyahu -UMD, H. Szu -SW Louisiana University, and C. Hsu -Trident Systems Inc.)

In this work, we concentrate on neural network classifiers and investigate how information obtained through a wavelet transform can be integrated in such a classifier. After a systematic dimensionality reduction by a Principal Component Analysis (PCA) technique, we apply a local spatial frequency analysis. This local analysis with a composite edge/texture wavelet transform provides statistical texture information of the Landsat imagery testset. The network is trained with both radiometric Landsat/Thematic Mapper (TM) bands and with the additional texture bands provided by the wavelet analysis.

The underlying assumptions that we are attempting to verify are that mixels (i.e., "border-line"/

mixed pixels, in a spatial/ spectral sense) contribute significantly to the overall misclassification of an image, and that (functions of) wavelet parameters will indicate how to single out these questionable pixels. Once detected, the classification of such pixels can be deferred to a post processing stage, at which other sophisticated schemes (e.g., relaxation-based) could be invoked, to yield an improved overall accuracy.

Results of this study are summarized in the 1995, 1996 and 1997 Annual Reports.

Parallel Implementations

(in collaboration with T. El-Ghazawi -GMU)

The pyramidal structure of the Mallat algorithm can be described as an iterative process containing the two major underlying operations of convolution and decimation. At each decomposition level, four new images are created, and as many layers as logN can be theoretically used for an NxN image. In general, for K decomposition levels, (3K + 1) sub-band images are produced.

The wavelet decomposition as well as the wavelet-based image registration algorithms were implemented on several parallel architectures, among which the MasPar MP-2, the Intel Paragon and the COTS (Commodity Off The Shelf) Beowulf architecture. Results and timings are reported in CESDIS Technical Reports 94-122, 94-125, and 97-203, as well as in the 1994, 1995, and 1997 Annual Reports.

GOES Follow-On AGSI Image Registration Subsystem

The previous results obtained in image registration and its implementations on different architectures were applied to the study of the image registration subsystem of the follow-on instrument to the GOES series, the AGSI (Advanced Geosynchronous Studies Imager). The two methods, edge-based and wavelet-based image registration, were chosen as potential methods to perform landmark registration and band-to-band co-registration. Computational requirements and trade studies are summarized in the two AGSI reports as well as in the 1997 Annual Report.

Publications

Stone, H. S., Le Moigne, J. and McGuire, M. Image Registration Using Wavelet Techniques, accepted for publication in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI.

Gosthtasby, A. and Le Moigne, J. (eds.) (1999). Special Issue on Image Registration. *Pattern Recognition Journal, 32*(1).

Le Moigne, J. and Cromp, R. F. (1999). Satellite Imaging and Sensing, in J.G. Webster (eds), *The Measurement, Instrumentation and Sensors Handbook* (CRC Press & IEEE Press), pp. 73.42 - 73.63.

Mount, D. M., Netanyahu, N. S. and Le Moigne, J. (1999). Efficient Algorithms for Robust Feature Matching. *Special Issue of Pattern Recognition on Image Registration, 32*(1), 17-38.

Le Moigne, J., Xia, W., Tilton, J. C., El-Ghazawi, T., Mareboyana, M., Netanyahu, N., Campbell, W. J. and Cromp, R. F. (1998). First Evaluation of Automatic Image Registration Methods. *IGARSS'98.*

Le Moigne, J. (1998). Image Registration Subsystem. *Report -AGSI Studies.*

An application of rotation- and translation-invariant overcomplete wavelets to the registration of remotely sensed imagery

Ilya Zavorine
University of Maryland, College Park
Department of Mathematics
(iaz@cs.umd.edu)

Abstract

In this work, wavelet coefficient maxima obtained from an orthogonal wavelet decomposition using Daubechies filters were utilized to register images in a multi-resolution fashion. Tested on several remote sensing datasets, this method gave very encouraging results. Despite the lack of translation-invariance of these filters, we showed that when using cross-correlation as a feature matching technique, features of size larger than twice the size of the filters are correctly registered by using the low-frequency subbands of the Daubechies wavelet decomposition. Nevertheless, high-frequency subbands are still sensitive to translation effects. In this work, we are considering a rotation- and translation-invariant representation developed by E. Simoncelli and integrate it in our image registration scheme. The two types of filters, Daubechies and Simoncelli filters, are then being compared from a registration point of view, utilizing synthetic data as well as data from the Landsat/ Thematic Mapper (TM) and from the NOAA Advanced Very High Resolution Radiometer (AVHRR).

1. Introduction

Automatic registration and resampling of remotely sensed data will be an essential element of future Earth satellite observation systems. New remote sensing systems will generate enormous amounts of data representing multiple observations of the same features at different times and/or by different sensors with, most often, these sensors being spread over multiple platforms. Automatic registration and resampling methods are indispensable for such tasks as data fusion, navigation, achieving super-resolution, or optimizing communication rates between spacecraft and ground systems. For all these tasks, accurate image registration is the first step, since a number of distortions prevent two images acquired either by the same sensor at different times or by two sensors at the same or different times from being "perfectly registered" to each other or to a fixed coordinate system. Distortions usually correspond to orbit and attitude anomalies, but some continuous nonlinear distortions are also due to altitude, velocity, yaw, pitch, and roll. It is very difficult to determine exact location within an image using only ancillary data and geo-location is usually computed by combining *navigation* and *registration*. Navigation corresponds to a "systematic correction" based on image acquisition models taking into account satellite orbit and attitude, sensor characteristics, platform/sensor relationship, Earth surface and terrain models and brings the registration accuracy within a few pixels. Image registration, on the other hand, corresponds to a "precision correction" based on landmarks and image features, and refines the geolocation to a subpixel accuracy. Registration is either applied after the navigation process, or both processes are integrated in a closed feedback loop. In this paper we will only consider the issue of feature-based, precision-correction automatic image registration.

In general, image registration can be defined as the process which determines the best match of two or more images acquired at the same or at different times by different or identical sensors. One set of data is taken as the *reference data*, and all other data, called *input data* (or *sensed data*), is matched relative to the reference data. The general process of image registration includes three main steps: (1) the extraction of features to be used in the matching process, (2) the feature matching strategy and metrics, and (3) the resampling of the data based on the correspondence computed from matched features. This paper only deals with steps (1) and (2). Currently, the most common approach to satellite image registration is to perform step (1) manually by interactive extraction of a few outstanding characteristics of the data, which are called *control points* (CP's), *tie-points*, or *reference points*. The CP's in both images (or image and map) are matched by pair and used to compute the parameters of a geometric transformation. But such a point selection represents a repetitive, labor- and time-intensive task which becomes prohibitive for large amounts of data, and often leads to large registration errors [1]. A large number of automatic image registration methods have been proposed and surveys can be found in [2-4]. Some of the features which are being utilized for step (1) are: original gray levels, edges, regions, and more recently wavelet features. According to [2], step (2) itself can be separated into:

- *the search space, i.e.* the class of potential transformations that establish the correspondence between input data and reference data. Transformations that are often used are rigid transformations (composed of a scaling, a translation and a rotation), affine transformations (composed of a rigid transformation, a shear and an aspect-ratio change; a shear in the x-axis transforms the x-coordinate into a linear combination of both x and y-coordinates, and the aspect-ratio is defined as the numerical ratio of image width to height), and polynomial transformations.

- *a search strategy*, which is used to choose which transformations have to be computed and evaluated. Local or global search, multi-resolution search or optimization techniques are examples of various search strategies.

- *a similarity metric*, which evaluates the match between input data and transformed reference data for a given transformation chosen in the search space. Correlation measurement has been the most often used but other methods such as a Hausdorff distance [5] can also be utilized.

A wavelet-based image registration approach has previously been proposed by the authors [4,6,7]. In this work wavelet coefficient maxima, obtained from an orthogonal wavelet decomposition using Daubechies filters [8], were utilized to register images in a multi-resolution fashion. Tested on several remote sensing datasets, this method gave very encouraging results. In the study reported in [9], we showed that when using cross-correlation as a feature matching technique, features of size larger than twice the size of the filters are correctly registered using the low-frequency subbands of the Daubechies wavelet decomposition. Nevertheless, features extracted from the high-frequency subbands are still sensitive to translation effects.

In this work, we are utilizing filters developed by E. Simoncelli [10-12] and we integrate them in our wavelet-based image registration scheme. The two types of filters, Daubechies and Simoncelli filters, are then being compared from a registration point of view, utilizing synthetic data as well as data from the Landsat/ Thematic Mapper (TM) and from the NOAA Advanced Very High Resolution Radiometer (AVHRR). Results are presented in section 3.

2. Some Background on Wavelet-based Image Registration of Satellite Imagery

2.a Previous Wavelet-Based Image Registration Method

Wavelet transforms provide a space-frequency representation of an image. In a wavelet representation, the original signal is filtered by the translations and the dilations of a basic function, called the "mother wavelet". In our wavelet-based registration, only discrete orthonormal bases of wavelets have been considered and are implemented by filtering, separately in rows and in columns, the original image by a high-pass and a low-pass filter, thus in a multi-resolution fashion [13]. At each level of decomposition, four new images are computed; each of these images is a quarter the size of the previous original image and represents the low frequency or high frequency information of the image in the horizontal and/or the vertical directions; images LL (Low/Low), LH (Low/High), HL(High/Low), and HH (High/High). Starting again from the "compressed" image (or image representing the low-frequency information, "LL"), the process can be iterated, thus building a hierarchy of lower and lower resolution images. Figure 1 summarizes the multi-resolution decomposition.

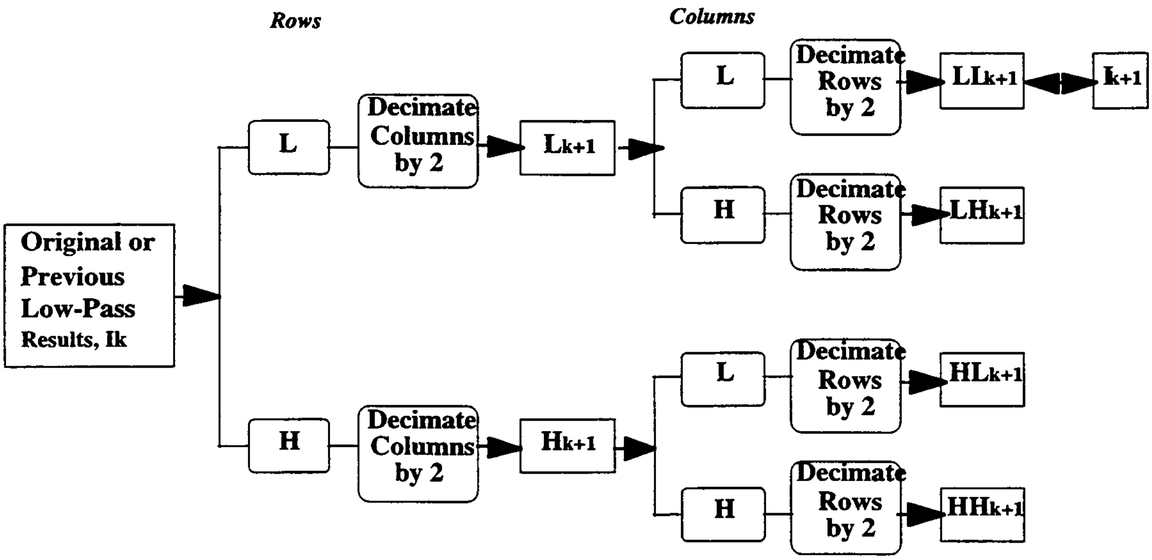


Figure 1: Decomposition by an Orthonormal Basis of Wavelets

Our wavelet-based method represents a three-step approach to automatic registration of remote sensing imagery. The first step involves the wavelet decomposition of the reference and input images to be registered. In the second step, we extract at each level of decomposition domain-independent features from both reference and input images. Finally, we utilize these features to compute the transformation function by following the multiresolution approach provided by the wavelet decomposition. Following the registration framework described in the previous section, our algorithm will utilize the four following components:

- *The feature space*
  Features are either chosen as the gray levels provided by the low-frequency LL compressed versions of the original image (for non-noisy images), or are based on the high-frequency information (e.g., maxima points of LH and HL images) extracted from the wavelet decomposition. In this second option, only those points whose intensities belong to the top x% of the histograms of these images are kept (x being a parameter of the program whose selection can be automatic); we call these points "maxima of the wavelet coefficients." These maxima are computed for all levels of the wavelet decomposition, for reference as well as input images.

- *The search space*
  In a first step, we assume the transformation to be either a rigid or an affine transformation. Both types of transformations include compositions of translations and rotations; therefore, as a preliminary study, our search space is composed of 2-D rotations and translations, and will be extended later to rigid and affine transformations. We look for rotations with angles included in the interval [0,90degrees] and for translations in the interval [0, half pixel-size of reference image].

- *The search strategy*
  Our search strategy follows the multi-resolution wavelet decomposition, iteratively from the deepest level of decomposition (where the image size is the smallest), until the first top level of decomposition, i.e. going from low resolution up to high resolution. For all levels of decomposition, the subband images of the reference image are successively transformed by all the transformations included in the search space. Then maxima of the transformed reference wavelet images and of the input wavelet images are extracted. The accuracy of this search increases when going from low resolution to high resolution. At each level the search focuses in an interval around the "best" transformation found at the previous level with an accuracy D and is refined at the next level up with an accuracy D/2. More details on this algorithm can be found in [4,6,7].

- *The similarity metric*
  At each level of decomposition and for each of the transformations, a correlation measure is computed between transformed reference maxima and input maxima. Another measure, based on a generalized Hausdorff distance, has also been studied and very encouraging results are reported in [5].

The previous algorithm is summarized in Figure 2. Tested on several datasets, the wavelet-correlation-based method described in this section performs with an accuracy of 1.66 pixels [14]. When using a statistically robust matching method based on a generalized Hausdorff distance, the first tests show that a subpixel accuracy can be obtained. More details on the results can be found in [4,14,15].

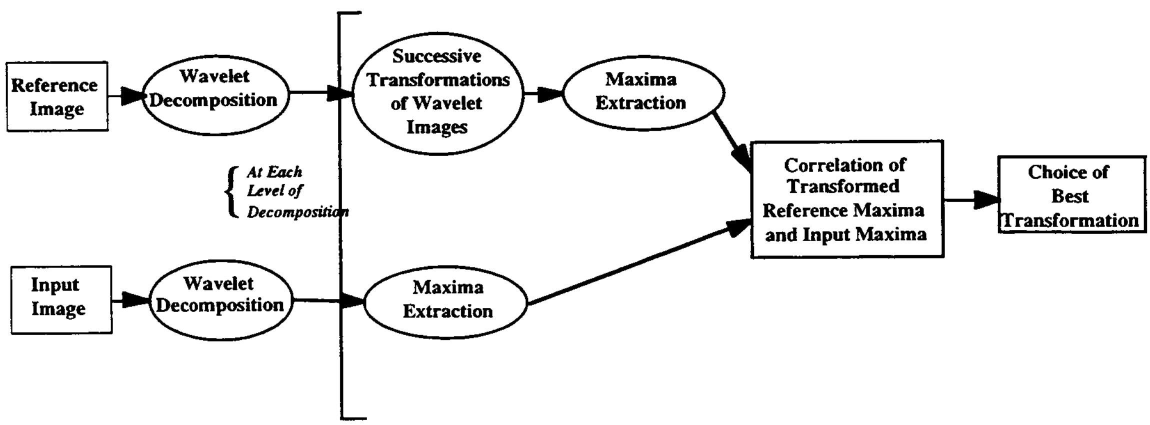


Figure 2: Summary of Our Wavelet-Correlation Image Registration Method

2.b Rotation and Translation Invariance Issues

According to the Nyquist criterion, in order to distinguish between all frequency components and to avoid aliasing, the signal must be sampled at least twice the frequency of the highest frequency component. Therefore, as pointed out in [10], "translation invariance cannot be expected in a system based on convolution and subsampling." When using a separable orthogonal wavelet transform (described in Figure 1), information about the signal changes within or across subbands. By lack of translation invariance, we mean that the wavelet transform does not commute with the translation operator, and similar remarks can be made relative to the rotation operator. Following these remarks, we conducted a study where the use of wavelet subbands is quantitatively assessed as a function of features' sizes. The study reported in [9] shows that when using cross-correlation, the method described in 2.a is still a useful registration scheme in spite of translation effects. The results are summarized here, see [9] for more details:

- the low-pass subband is relatively insensitive to translation, provided that the features of interest have an extent at least twice the size of the wavelet filters.
- the high-pass subband is more sensitive to translation, but the peak correlations are still high enough to be useful.

Following this study, the work presented in this paper only considers the high-pass subbands and look at rotation- and translation-invariant filters [10] in order to create the feature space. Although the scheme described in Figure 2 would be more optimal if a different similarity metrics were used, we keep the same correlation framework for the only purpose of comparing Daubechies and Simoncelli's filters under the same conditions and for registration purposes. Experiments involving a different search strategy and different similarity metrics are currently being performed as a continuation of this work.

3. Use of a Rotation and Translation-Invariant Representation for Image Registration

3.a Rotation- and Translation-Invariant Representation

The method described by E. Simoncelli in [10-12] enables the construction of translation- and rotation-invariant filters by relaxing the critical sampling condition of the wavelet transforms. By invariance, it is meant that the information contained in a given subband will be invariant to translation or rotation. The resulting representation is equivalent to an overcomplete wavelet transform; it is not an orthogonal representation but is an approximation of a "tight-frame" [8], i.e. invertible. The Steerable Pyramid described in [10] is summarized in Figure 3, where only the analysis decomposition is shown. H0 is a high-pass filter, L0 and L1 are two low-pass filters, and {B0, ..., Bk} represents a set of oriented band-pass filters which ensure the representation to be rotation-invariant. In order to ensure translation-invariance, the output of the high-pass filter and of the band-pass filters are not subsampled. In addition, the portion of the signal which is iteratively decomposed by the band-pass and the low-pass filters does not contain the larger high frequency components and has been preprocessed by the low-pass filter, L0, thus removing most aliased components.

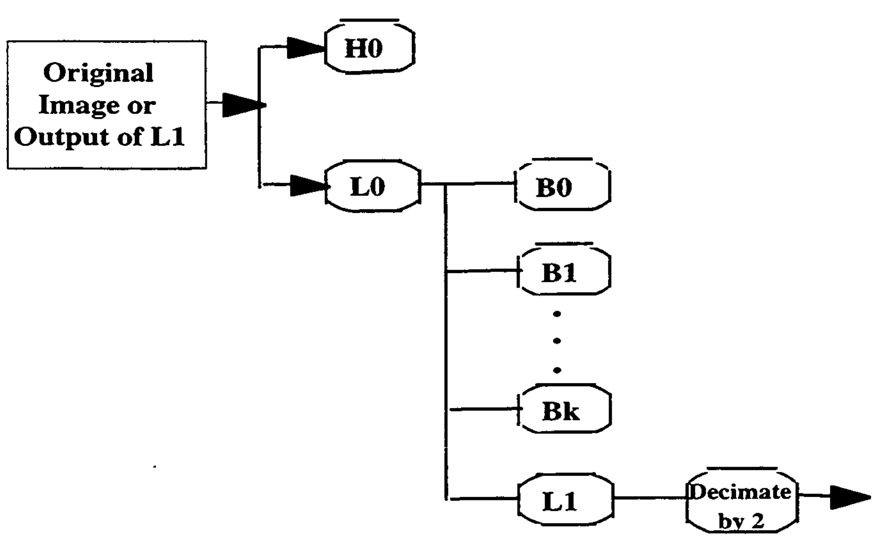


Figure 3: Decomposition by a Steerable Pyramid

As stated in [10], this representation is overcomplete by a factor of 4k/3, where k is the number of oriented band-pass filters. In the experiments shown below, in order to optimize the computational speed, we chose k=1. The decomposition was iterated 3 times and the subbands which were considered for feature selection in the registration algorithm are {S0,S1,S2,S3} as shown in Figure 4.

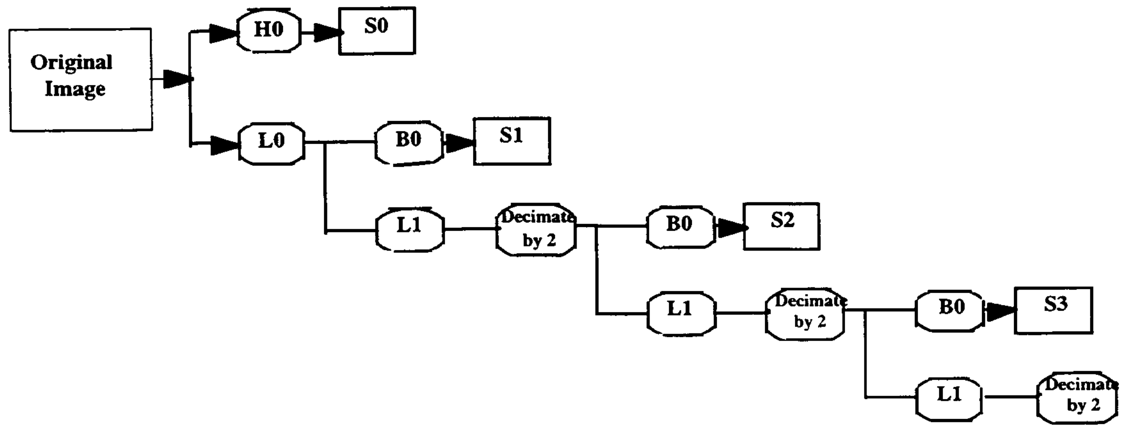


Figure 4: 3-Level Decomposition by a Steerable Pyramid Using Only 1 Oriented Band-Pass Filter

3.b Results of the Comparative Study

3.b.1 Description of the Parameters

As we previously stated, the purpose of this study is to vary the type of features used in the registration process described in section 2a and in Figure 2 and observe the results when tested on multiple datasets.

Using the Daubechies filters and the separable orthogonal decomposition of Figure 1, three levels of decomposition are processed and the feature space is composed of the maxima of images {LH2,HL2}, {LH1,HL1}, and {LH0,HL0} for each respective refinement iteration. These images correspond to a decimation by 8, 4, and 2 of the original image, respectively.

Using the Simoncelli filters and the Steerable Pyramid decomposition of Figure 4, three levels of decomposition are processed and the feature space is composed of the maxima of images {S3},{S2}, and {S1}. These images correspond to a decimation of 8, 4, and 2 of the original image, respectively. Although using the features provided by S0 would significantly improve the results (since it is of size identical to the original's), this information has been purposely left out in order to provide a consistent comparison of the results between the two types of filters.

Since, at each level, Daubechies' features are obtained from two different subbands and Simoncelli's features from only one subband, the maxima extraction threshold has been tested at {15%,20%,25%} for Daubechies' features and at {30%,40%,50%}, respectively, for Simoncelli's features, thus keeping the same number of features to be correlated in both cases.

3.b.2 Description of the Test Datasets

The tests were performed on eight different datasets, which are also illustrated in Figures 5 to 9.

Dataset #1, "SYNTH". A 512x512 synthetic image formed of various geometric shapes was created. Used as a reference image, transformed images are generated by applying compositions of rotations R(Θ) and T=(Tx,Ty) where Θ=(0,5,10,15,20,25} degrees and Tx,Ty={5,10,15,20} pixels. This results in a dataset of 102 images including the reference image.

Datasets #2-#5, "NSYNTH.G2", "NSYNTH.G5", "NSYNTH.G10", "NSYNTH.G20". The previous dataset was altered by white Gaussian noise of variance 2,5,10, and 20, respectively.

Dataset #6, "GIRL". The reference image for this dataset is a 512x512 digitized photograph of a face containing very little noise. The transformations of the reference image include the set of translations {(6,4),(8,10),(12,8),(14,12),(16,20),(20,60)}, rotations of angles {5,10,15,20,25} degrees, and the same rotations combined with the translations {(2,4),(6,4),(20,60)}. This dataset contains 27 files.

Dataset #7, "TM". The reference TM image is a 512x512 portion of Band 2 of a Landsat-Thematic Mapper (TM) scene over the Pacific Northwest. Transformations are identical to the ones described for dataset #6, with 27 files.

Dataset #8, "AVHRR". This dataset is much smaller but represents a "real-life dataset". It is a series of thirteen 512 rows by 1024 columns AVHRR/LAC images over South Africa. Raw AVHRR data are navigated and georeferenced to a geographic grid that extends from -30.20 S, 15.39 E (upper left) -34.79 S, 24.59 E (lower right). The navigation process uses an orbital model developed at the University of Colorado [16] and assumes a mean attitude behavior (roll, pitch and yaw) derived using Ground Control Points [17]. A map of the coastline derived from the Digital Chart of the World (DCW) is generated for the same geographic grid and is also available for this dataset. In this case, the actual transformation is not known, but results of a manual registration are used to assess the accuracy of the automatic registration.

3.b.3 Results

Registration results obtained with the two different types of filters are summarized in Tables 1 and 2 and graphically represented in Figure 10. These results show that the two types of filters give similar results for ideal or low-noise images but as soon as the noise level increases, the registration accuracy is much more stable with Simoncelli's filters. Overall the translation accuracy obtained with these filters stays around 1 pixel, and even reaches subpixel accuracy for the "GIRL" and "TM" datasets; while the accuracy using Daubechies' filters greatly varies depending on the contents of the images. The results are consistent for all levels of thresholds chosen in the maxima selection, even when the noise level increases.

4. Conclusion and Future Work

This paper presented an image registration method based on wavelets and overcomplete wavelets. It was shown that, as expected and due to their translation- and rotation- invariance, Simoncelli's filters perform better than Daubechies' filters. Quantitative measurements support this conclusion.

As we recognize that the exhaustive search involving multiple cross-correlations is not optimal, we are looking at other search strategies and similarity metrics, such as optimization and robust matching. Future work will also exploit the full rotation-invariance capability of the steerable filters by varying the number of band-pass filters.

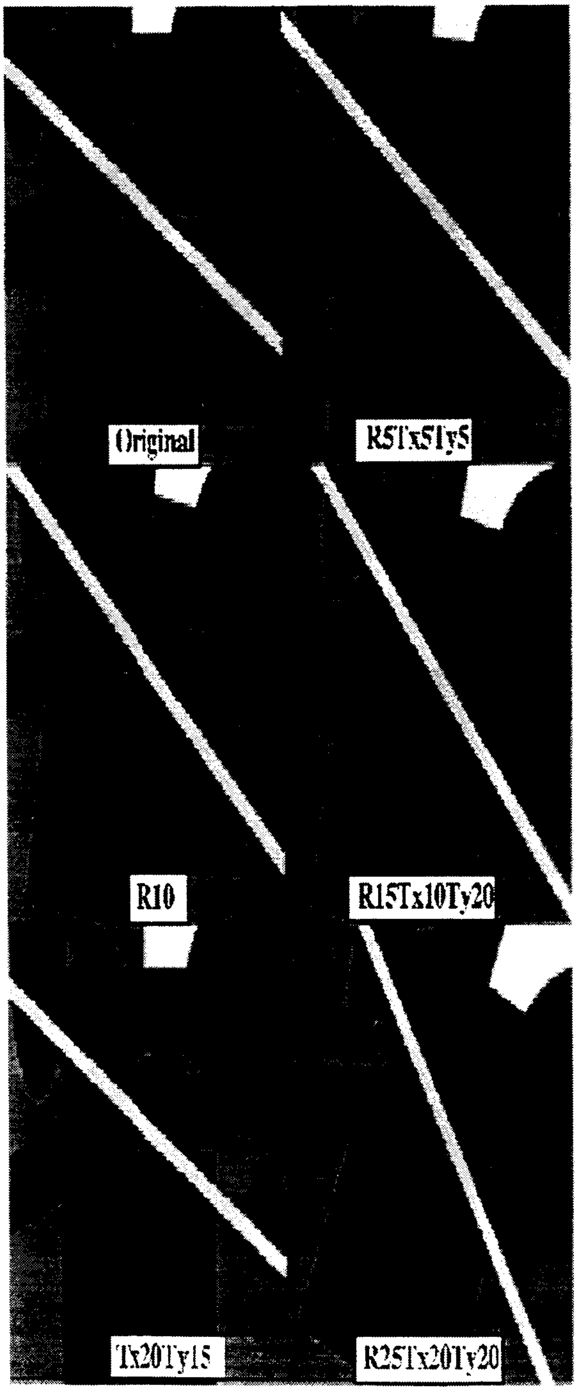


Figure 5: Dataset #1 ("SYNTH") Reference and Some Transformations

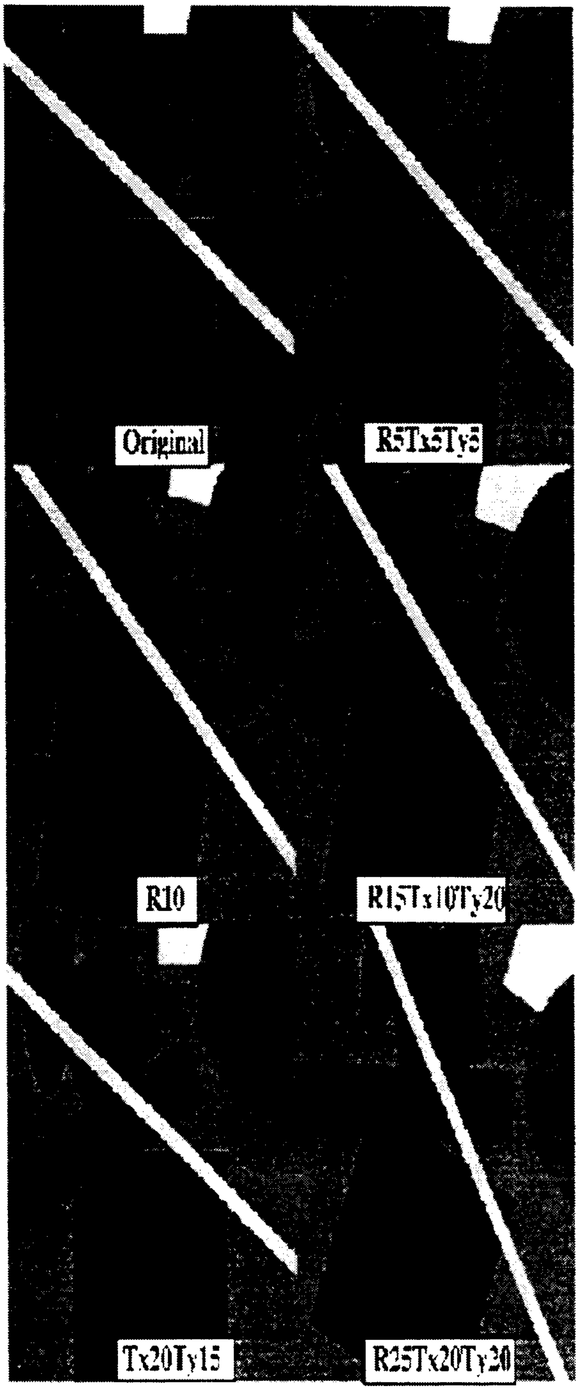


Figure 6: Dataset #5 ("NSYNTH.G20") Reference and Some Transformations

Figure 7: Dataset #6 ("GIRL") Reference and Some Transformations

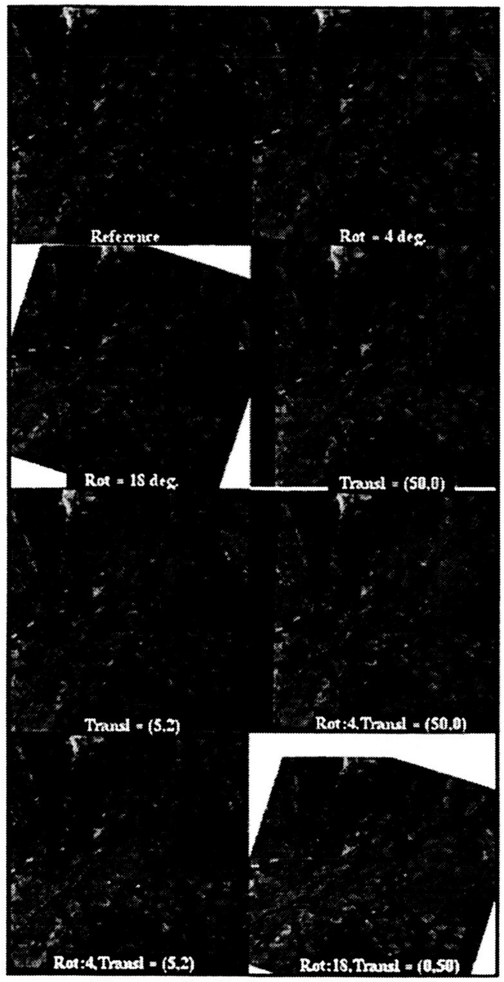


Figure 8: Dataset #7 ("TM") Reference and Some Transformations

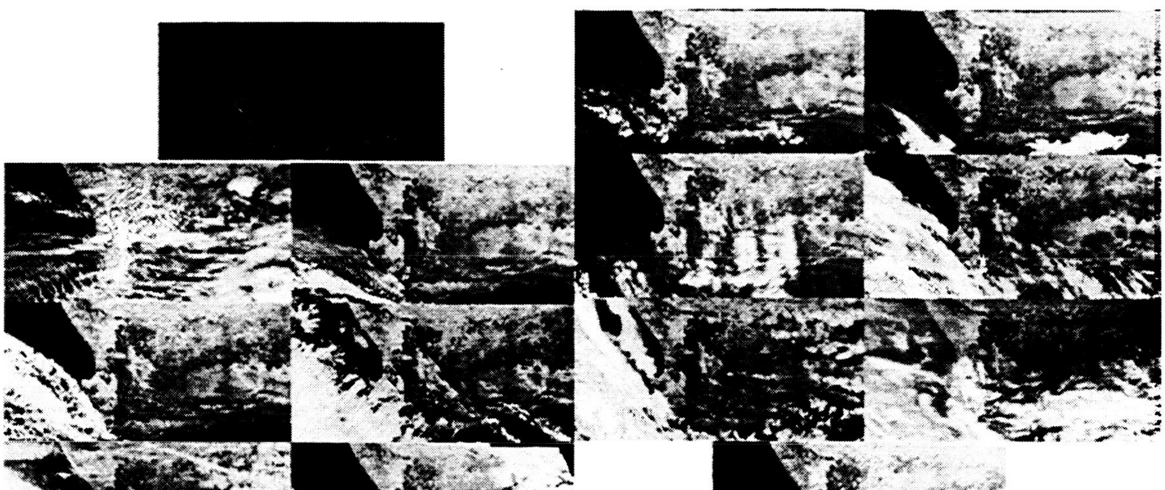


Figure 9: Dataset #8 ("AVHRR") Series of Thirteen Multi-Temporal AVHRR Images

Acknowledgments

The authors wish to acknowledge Dr. Milton Halem for his support to this work, as well as the continuing support of the Applied Information Sciences Branch (NASA/Code 935) and of the Center of Excellence in Space and Data Information Sciences (CESDIS). The authors would also like to thank Pr. Eero Simoncelli for a very useful discussion about his research and his useful comments about our work.

References

Townshend, J. R. G. and Justice, C. O. (1992). The Impact of Misregistration on Change Detection. *IEEE Transactions on Geoscience and Remote Sensing, 30*(5).

Brown, L. (1992). A Survey of Image Registration Techniques. *ACM Computing Surveys, 24*(4).

Fonseca, L. M. G. and Manjunath, B. S. (1996). Registration Techniques for Multisensor Sensed Imagery. *Photogrammetric Engineering and Remote Sensing Journal, 62* (9), 1049-1056.

Le Moigne, J., Campbell, W. J. and Cromp, R. F. An Automated Parallel Image Registration Technique of Multiple Source Remote Sensing Data. CESDIS TR-96-182 - submitted to *IEEE Transactions in Geoscience and Remote Sensing.*

Mount, D. M., Netanyahu, N. S. and Le Moigne, J. (1999). Efficient Algorithms for Robust Feature Matching, to appear in *Special Issue of Pattern Recognition on Image Registration.*

Le Moigne, J. Parallel Registration of Multi-Sensor Remotely Sensed Imagery Using Wavelet Coefficients. *SPIE's OE/Aerospace Sensing, Wavelet Applications Conference*, 432-443.

Le Moigne, J. Towards a Parallel Registration of Multiple Resolution Remote Sensing Data. *IGARSS'95, International Geoscience and Remote Sensing Symposium*, Firenze, Italy, pp.1011-1013.

Daubechies, I. (1991). Ten Lessons on Wavelets. *CMBS-NSF Series Applications in Mathematics, SIAM.*

Stone, H, Le Moigne, J. and McGuire, M. (1997). Registration Using Wavelet Techniques. *26-th Applied Imagery Pattern Recognition Workshop (AIPR'97)*, Washington DC, submitted to *IEEE-PAMI.*

Simoncelli, E., Freeman, W., Adelson, E. and Heeger, D. (1992). Shiftable Multiscale Transforms. *IEEE Transactions on Information Theory, 38*(2).

Simoncelli, E. and Freeman, W. (1995). The Steerable Pyramid: A Flexible Architecture for Multi-Scale Derivative Computation. *2-nd Annual IEEE International Conference on Image Processing*, Washington, DC.

Karasaridis, A. and Simoncelli, E. A Filter Design Technique for Steerable Pyramid Image Transforms. *Proceedings ICASSP-96*, Atlanta, GA.

Mallat, S. (1989). A Theory for Multiresolution Signal Decomposition. *IEEE Pattern Analysis and Machine Intelligence, PAMI-11*(7).

Le Moigne, J., Xia, W., Tilton, J. C., El-Ghazawi, T., Mareboyana, M., Netanyahu, N., Campbell, W. J. and Cromp, R. F. (1998). First Evaluation of Automatic Image Registration Methods. *IGARSS'98.*

El-Ghazawi, T., Chalermwat, P. and Le Moigne, J. (1997). Wavelet-Based Image Registration on Parallel Computers. *SuperComputing'97*, San Jose.

Rosborough, G. W., Baldwin, D.G., Emery, W. J. (1994). Precise AVHRR Image Navigation. *IEEE Transactions on Geoscience and Remote Sensing, 32*(3).

Baldwin, D., Emery, W. (1995). Spacecraft Attitude Variations of NOAA-11 Inferred from AVHRR Imagery. *International Journal of Remote Sensing, 16*(3), 531-548.

SYNTHETIC	Threshold 15		Threshold 20		Threshold 25	
	Rotation	Translation	Rotation	Translation	Rotation	Translation
Orthogonal/Daubechies						
Accuracy	0.0000	1.0990	0.0000	1.0990	0.0000	1.0990
Standard Deviation	0.0000	0.7513	0.0000	0.7513	0.0000	0.7513
Steerable/Simoncelli						
Accuracy	0.0000	1.1089	0.0000	1.1287	0.0000	1.1287
Standard Deviation	0.0000	0.9109	0.0000	0.9299	0.0000	0.9299
SYNTHETIC-NOISE 2	**Threshold 15**		**Threshold 20**		**Threshold 25**	
	Rotation	Translation	Rotation	Translation	Rotation	Translation
Orthogonal/Daubechies						
Accuracy	0.0000	1.5842	0.0000	1.4752	0.0000	1.3366
Standard Deviation	0.0000	1.2127	0.0000	1.1739	0.0000	0.9978
Steerable/Simoncelli						
Accuracy	0.0000	1.2277	0.0000	1.1089	0.0000	1.1287
Standard Deviation	0.0000	1.2733	0.0000	0.9109	0.0000	0.9299
SYNTHETIC-NOISE 5	**Threshold 15**		**Threshold 20**		**Threshold 25**	
	Rotation	Translation	Rotation	Translation	Rotation	Translation
Orthogonal/Daubechies						
Accuracy	0.0000	1.6535	0.0000	1.4752	0.0000	1.3564
Standard Deviation	0.0000	1.4032	0.0000	1.2396	0.0000	1.0301
Steerable/Simoncelli						
Accuracy	0.0000	1.3663	0.0000	1.3267	0.0000	1.1683
Standard Deviation	0.0000	1.5397	0.0000	1.3867	0.0000	1.0056
SYNTHETIC-NOISE 10	**Threshold 15**		**Threshold 20**		**Threshold 25**	
	Rotation	Translation	Rotation	Translation	Rotation	Translation
Orthogonal/Daubechies						
Accuracy	0.1188	3.5248	0.0792	2.9802	0.0792	2.1188
Standard Deviation	0.4288	2.4600	0.3045	2.4655	0.3045	1.6488
Steerable/Simoncelli						
Accuracy	0.0000	1.4257	0.0000	1.3663	0.0000	1.3267
Standard Deviation	0.0000	1.5113	0.0000	1.4331	0.0000	1.3867
SYNTHETIC-NOISE 20	**Threshold 15**		**Threshold 20**		**Threshold 25**	
	Rotation	Translation	Rotation	Translation	Rotation	Translation
Orthogonal/Daubechies						
Accuracy	3.2970	13.3663	3.0693	12.5446	2.9208	12.0297
Standard Deviation	2.9610	4.5917	2.0646	5.0942	2.0904	6.0419
Steerable/Simoncelli						
Accuracy	0.0000	2.1386	0.0000	1.6436	0.0000	1.5842
Standard Deviation	0.0000	2.1937	0.0000	1.7439	0.0000	1.6245

Table 1: Registration Results for Both Types of Filters on Synthetic Datasets #1-#5

GIRL	Threshold 15 Rotation	Translation	Threshold 20 Rotation	Translation	Threshold 25 Rotation	Translation
Orthogonal/Daubechies						
Accuracy	0.5556	3.8889	0.2963	2.3333	0.1481	1.7037
Standard Deviation	1.6178	2.8458	0.5316	1.8459	0.3552	1.3282
Steerable/Simoncelli						
Accuracy	0.0000	0.4815	0.1852	1.0741	0.1852	0.9259
Standard Deviation	0.0000	0.4997	0.4743	1.7832	0.4743	1.3032

TM	Threshold 15 Rotation	Translation	Threshold 20 Rotation	Translation	Threshold 25 Rotation	Translation
Orthogonal/Daubechies						
Accuracy	1.9091	7.0606	0.2727	2.5455	0.2727	2.2121
Standard Deviation	2.8000	7.2360	0.9301	3.7424	0.9621	3.1019
Steerable/Simoncelli						
Accuracy	0.0000	0.5455	0.0000	0.5455	0.0000	0.5455
Standard Deviation	0.0000	0.4979	0.0000	0.4979	0.0000	0.4979

AVHRR	Threshold 15 Rotation	Translation	Threshold 20 Rotation	Translation	Threshold 25 Rotation	Translation
Orthogonal/Daubechies						
Accuracy	3.2500	12.0000	3.0833	9.9167	3.8333	9.0000
Standard Deviation	0.3727	2.4267	0.2764	5.8878	0.0000	2.4944
Steerable/Simoncelli						
Accuracy	0.1667	2.3333	0.0833	4.0000	0.0000	2.6667
Standard Deviation	1.7381	5.4620	2.5644	6.2244	3.3624	6.7330

ALL SETS	Threshold 15 Rotation	Translation	Threshold 20 Rotation	Translation	Threshold 25 Rotation	Translation
Orthogonal/Daubechies						
Accuracy	0.0588	1.5650	0.1173	1.6212	0.1098	1.2661
Standard Deviation	1.7035	5.4647	1.5500	5.0611	1.5770	5.0474
Steerable/Simoncelli						
Accuracy	0.0035	1.3744	0.0104	1.3154	0.0087	1.2392
Standard Deviation	0.0588	1.5650	0.1173	1.6212	0.1098	1.2661

Table 2: Registration Results for Both Types of Filters on Datasets #6-#8 and Overall Results

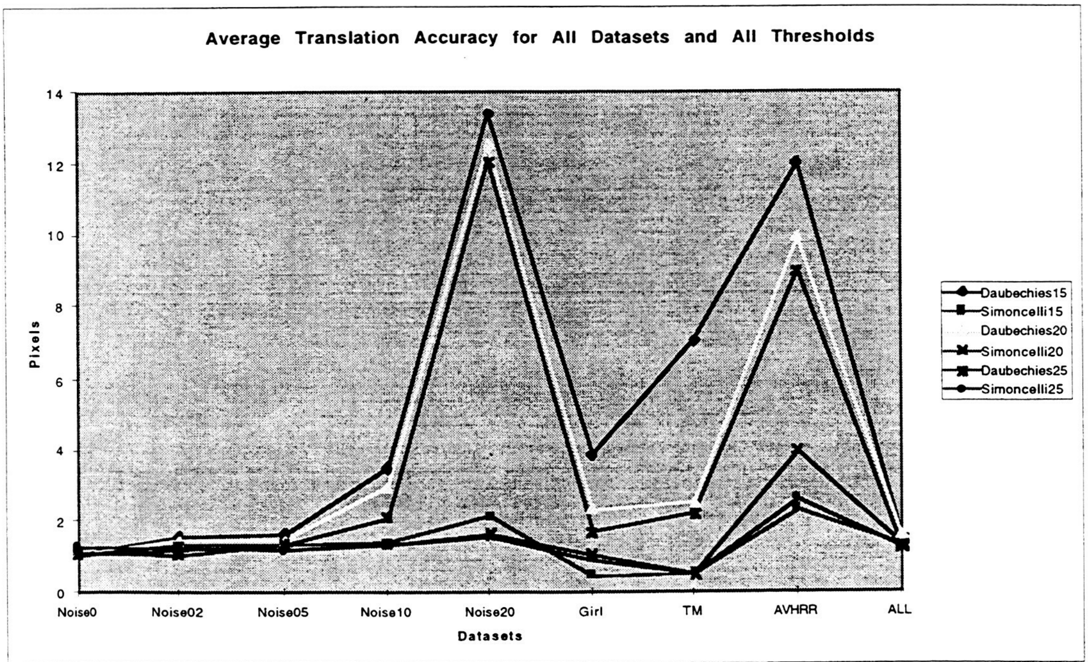


Figure 10: Summary of Results for Translation Accuracy

COMPUTATIONAL RESEARCH TEAM

Richard G. Lyon
University of Maryland Baltimore County
Department of Electrical Engineering
(lyon@jansky.gsfc.nasa.gov)

Timothy P. Murphy
University of Maryland Baltimore County
Department of Computer Science and Electrical Engineering
(murphy@albert.gsfc.nasa.gov)

Miodrag Rancic
University of Maryland Baltimore County
Department of Computer Science and Electrical Engineering
(mrancic@ciga.gsfc.nasa.gov)

George Lake
University of Washington
Department of Astronomy
(lake@astro.washington.edu)

Adam Frank
University of Rochester
Department of Physics and Astronomy
(afrank@alethea.pas.rochester.edu)

Phase Retrieval Based Coronagraphy for Extra-Solar Planetary Detection

Richard G. Lyon
University of Maryland Baltimore County
Department of Electrical Engineering
(lyon@jansky.gsfc.nasa.gov)

Profile

Richard G. Lyon is a Research Scientist at the Center of Excellence in Space Data and Information Sciences, from the University of Maryland Baltimore County, and is associated with the Optical Systems and Characterization Project (OSCAR) at NASA/Goddard Space Flight Center. OSCAR is currently funded by both NASA/GSFC and NASA/JPL to conduct research into computational and hardware methods of optimal information extraction, wavefront sensing and imaging science. Mr. Lyon holds 4 NASA awards for his work on the Hubble Space Telescope and on the Advanced Geosynchronous Studies (AGS) Imager. Mr. Lyon is also Co-I on the pre-Phase A Integrated Instrument Science Module concept study for NGST to design a coronagraphic instrument for NGST. Mr. Lyon holds a Bachelor of Science in Physics from the University of Massachusetts and a Master of Science in Optics from the University of Rochester with work towards a Ph.D. in Optics at the University of Rochester. He is a member of the Optical Society of America (OSA) and the Society for Photo-Instrumentation Engineers (SPIE) and

From 1987 to 1992 Mr. Lyon was employed by Hughes Danbury Optical Systems (now Raytheon Optical Systems Inc.) as an optical systems engineer in the Space Sciences directorate. In that capacity he served as principal investigator of Hubble Space Telescope phase retrieval efforts to determine the on-orbit telescope error. From January 1993 to June 1994 Mr. Lyon worked as a research analyst for Radex Incorporated where he designed, developed and implemented automated celestial image processing algorithms for the Mid-Course Space Experiment (MSX), a U.S. Air Force radiometric satellite. In June 1994 he became a principal engineer with Hughes STX where he conducted research into the design and development of optical and image processing algorithms to operate in massively parallel computational environment, including image restoration and image deconvolution algorithms for the Hubble Space Telescope.

Mr. Lyon is currently associated with the Optical Systems Characterization and Analysis Research (OSCAR) project [1][]2] has been actively researching optimal information extraction problems including:

- Phase Retrieval based Coronagraph for the Next Generation Space Telescope (NGST). This is discussed in more detail in the report below.

- On-Board active optical control loop, based on phase retrieval, for the Next Generation Space Telescope and for the Developmental Comparative Active Telescope Testbed (DCATT). This is discussed in detail in [3].

- Phase Diverse wavefront sensing for the NASA EO-3/RedEye mission. This is discussed in detail in [4].

The common thread of this research has been the marriage of computing with optics for optimal information extraction.

Phase Retrieval based Coronagraph for NGST

NASA/JPL, in collaboration with author, proposed and was funded to study the use of phase retrieval for a coronagraph on the Next Generation Space Telescope [5]. Future potential NASA space-based astronomical missions such as the Next Generation Space Telescope (NGST), and others, will open the exciting possibility of direct images of planets around nearby stars. A bright central stellar source is typically 6 to 9 orders of magnitude brighter than the planet. In most cases the planets light is lost in the diffraction and coherent scattering halo, commonly called "speckle", from the bright central star. The diffraction is caused by the optical wavefront propagating through the finite size telescope optics. The larger the telescope the smaller the diffraction effects. The speckle is caused the spatially coherent interference between the residual structure on the optical surfaces. The smoother the optics the less the speckle. The simplest approach to minimizing the diffraction and speckle is to have a very large, very smooth surface telescope in space. A very expensive proposition. However a well designed coronagraph reduces the diffraction glare and phase retrieval allows us to determine and remove the speckle. A coronagraph is essentially an instrument with series of apertures and masks which tailor the shape of the telescope diffraction. Use of phase retrieval with an optical control loop allows for removal of the speckle. Phase retrieval is a computationally intensive algorithm.

Figure 1 shows a simplified schematic of an optical coronagraph. The light propagates from left to right. The top row of figure 1 shows a plane wave incident on the telescope pupil and is focused to the first focal plane, known as the occulting plane. The bottom row shows the optical intensity in each of the respective planes. An apodized mask is inserted which blocks part of the light from the bright central source. This mask only removes the central core but leaves the diffraction halo as seen in the lower row. The wavefront then propagates to the Lyot plane containing another mask. The effect of the mask in the occulting plane is to introduce a bright ring in the Lyot plane. It is just this ring which is masked in the Lyot plane. Finally the image is brought to focus. Various size occulting and Lyot masks are possible depending on wavelength and how far the planet is from the central source.

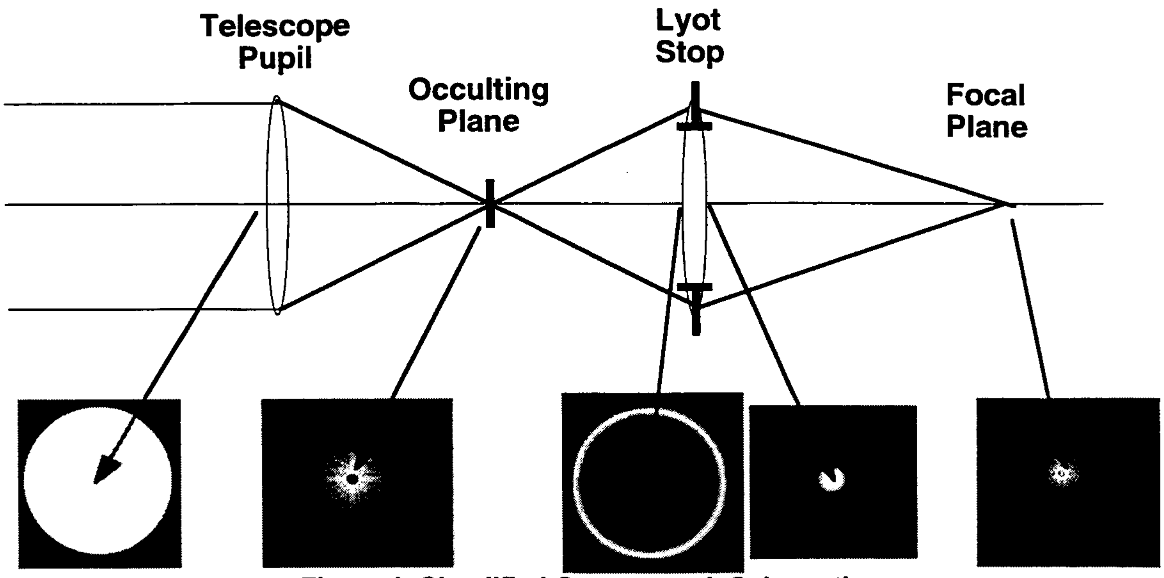


Figure 1: Simplified Coronagraph Schematic

The image in the final focal plane has been significantly attenuated from the image in the occulting plane. This is not evident in Figure 1 since the color table has been stretched to show the details. Figure 2 shows this attenuation. Plotted is the radial intensity versus distance from the center, both axes are logarithmic (base 10). The top trace shows the image with no occulting or Lyot mask. The bottom 4 traces show the image with a hard edged 10 ring occulting mask and for increasing levels of Lyot apodization, from 0 to 30%, i.e., the Lyot pupil function is 30% smaller. One can easily see that theoretically 5 orders of magnitude of attenuation are possible outside of 20 rings. Thus, we should be able to detect planets with this configuration. Greater attenuations are possible by "apodization", i.e., softening the edges of both the occulting and Lyot masks.

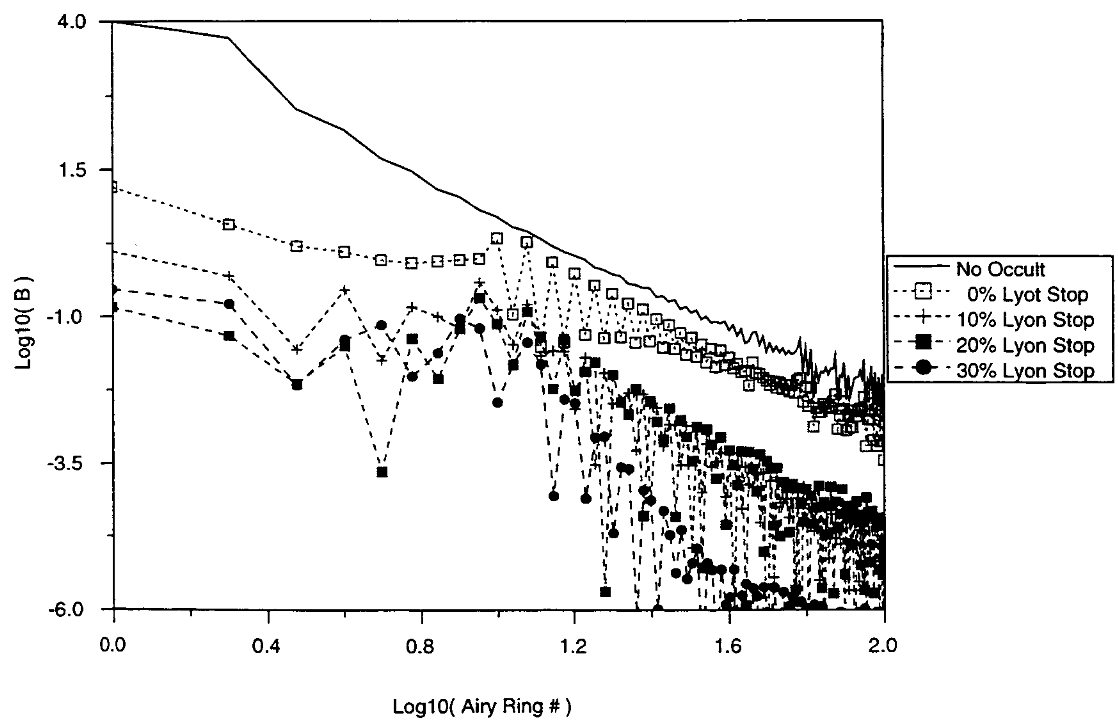


Figure 2: Reduction in Diffraction due to Coronagraph

Figure 2 shows graphically how the coronagraph works but neglects the effects of speckle. Figure 3 shows the speckle. The left of Figure 3 is a speckle free image of the final focal plane and the right image shows the image with speckle. It would be hard to extract a planet from the right side of Figure 3, since it would appear much like a speckle.

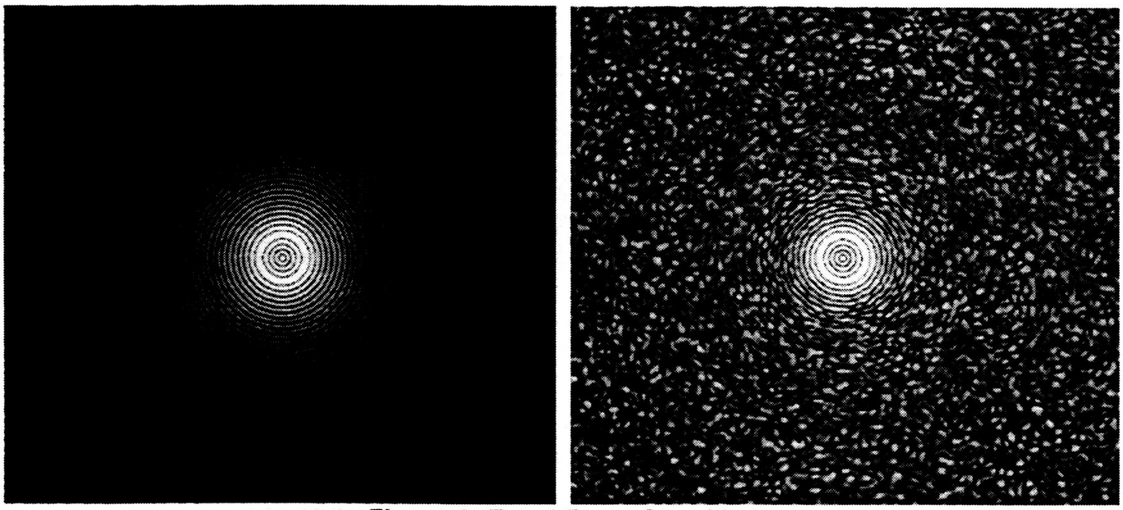
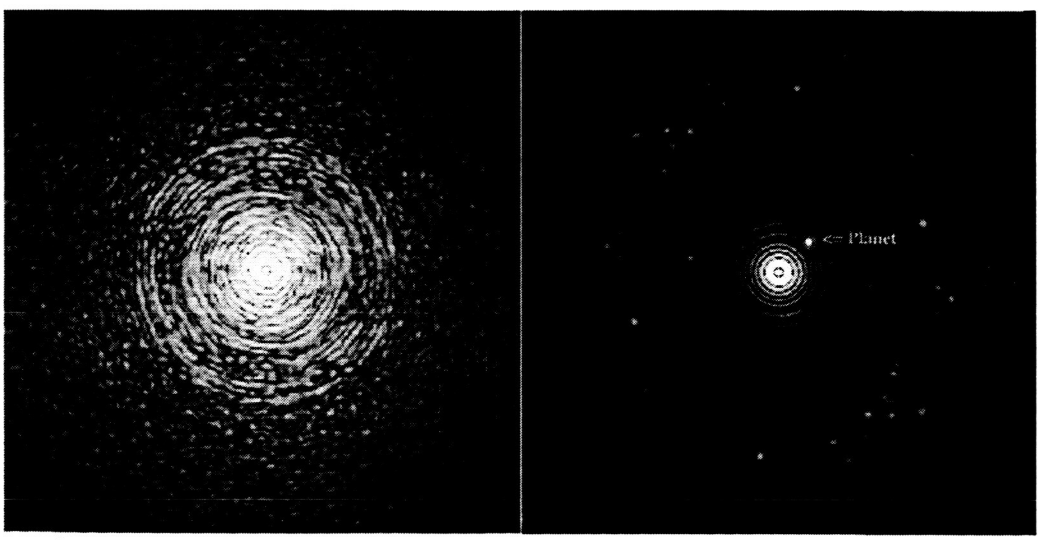


Figure 3: Focal Plane Speckle

Figure 4 shows a more realistic simulation of both mid spatial frequency polish structure on the mirror surfaces and the effects of focal plane speckle. The leftside of Figure 4 is a final focal plane image after passing through the coronagraph. The diffraction from the residual surface polish marks are clearly visible as a series of random nearly concentric rings about the core. The speckle is also visible as random spots decaying in intensity away from the core. A planet is hidden in the left side of Figure 4. The rightside is after phase retrieval has been used to determine the mid- and high-spatial frequency phase and subsequent correction by an active optical control loop. The planet can now be clearly seen to the upper right of the core. For more detail on the phase retrieval process see [2][6] and the references therein.



Lyot type Coronagraph | **Lyot type Coronagraph with phase retrieval** (planet at 8 λ/D)

Figure 4: Lyot Coronagraph with Phase Retrieval

References

Lyon, R. G., Dorband J. E., Hollis, J. M., and Murphy, T. P. (1999). Optical Modeling Helps Telescopes see more clearly. *Insights Magazine, High Performance Computing and Communications*, ed. Jarrett Cohen

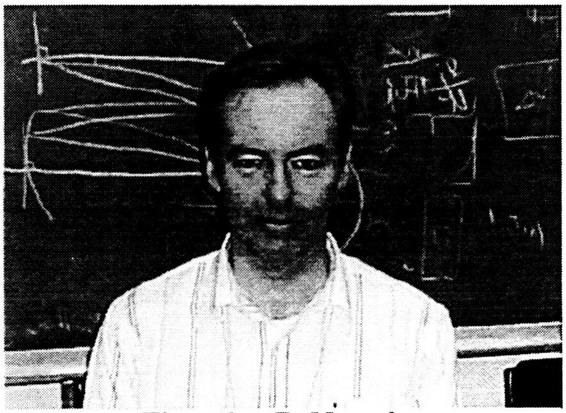
Lyon, R. G., Hollis, J. M, Dorband, J., Murphy, T. P. (1998). Extrapolating HST Lessons to NGST, Invited Paper, *Optics and Photonics News*, *9*(7).

Lyon, R. G., Murphy T. P., Dorband J. E., Young, E. W., Wilson, M. E., and Davila, P. S. (1999). DCATT Wavefront Sensing and Optical Control Study. *Developmental Comparative Active Telescope Testbed Report # WFSC-0001*, NASA/Goddard Space Flight Center.

Lyon, R. G., Murphy, T. P., Feinberg, L. D., and Jenstrom, D. T. (1999). Phase Diversity Wavefront Sensing for the NASA/EO-3 RedEye Camera. *1999 OSA Annual Meeting*, Santa Clara, CA.

Trauger (JPL), Backman, D. E. (Franklin & Marshall), Beichman, C. A., Grillmair, C. J. (IPAC), Djorgovski, S. G., Kulkarni, S. R., Martin, C. D. (Caltech), Gaidos, E., Gursel, Y., Macenka, S., Ressler, M. E., Sahai, R., Shao, M., Stapelfeldt, K. R., Werner, M. W., Lyon, R., and Ealey, M. (Xinetics) (1999). Strategies for High-Contrast Imaging with NGST. *193 AAS Meeting*, Austin Tx.

Lyon, R. G., Dorband, J. E., and Hollis, J. M. (1997). Hubble Space Telescope Faint Object Camera Calculated Point Spread Functions. *Applied Optics, 36*(8).



Timothy P. Murphy
University of Maryland Baltimore County
Department of Computer Science and Electrical Engineering
(murphy@albert.gsfc.nasa.gov)

Profile

Mr. Murphy holds a B.S. and M.S. in electrical engineering from Columbia University. He was employed by the Perkin-Elmer Corporation from 1980 to 1991 where his responsibilities included analysis of space radiation effects, design of radiation shielding for electro-optical sensors, research into DSP applications for spectroscopy instruments and electro-optical system analysis. He was also employed by Dalsa, Inc., Waterloo ONT in 1991 and 1992 for design of a data acquisition and test station for the first 25 million-pixel CCD. From 1993 to 1998 he was employed by GN Nettest, Fiber Optics Division, Utica, N.Y., where he designed signal processing software for fiber optic equipment.

At CESDIS Mr. Murphy is working on optical control software for the NASA High Performance Computing and Communication (HPCC) Remote Exploration and Experimentation (REE) program and laboratory research in phase diverse imaging.

Optical Alignment and Control for NGST: Prototype Application for Fault Tolerant Computing in Space.

As part of NASA's REE program, we have supplied prototype scalable, multiprocessor computer applications for Optical Alignment and control of the Next Generation Space Telescope (NGST)[1]. Currently, the applications run on a Linux-Beowulf operating system. They will be ported to the REE fault tolerant computing platform as it becomes available.

The three parts of this application delivered to date are:

- Phase Retrieval by Misell Algorithm
- Phase Unwrapping by directed acyclic graph
- Actuator Fitting by least-squared wavefront error

Background

The NGST will be an infrared telescope with a segmented 8-meter primary mirror. Its mission orbit will most likely be at the L2 Lagrange point, some 1.5 million kilometers from earth. Putting such a large structure at a remote location necessitates saving weight in every possible way. The telescope design is based on the principle that the telescope optics will be very light weight. They will be so light that they will be somewhat flexible and prone to optical aberrations. These aberrations will be corrected through the use of an active optical system [2]. The active optical system consists of actuators to move the secondary mirror (SM) and primary mirror (PM) and also a deformable mirror (DM). The DM will have perhaps 349 actuators to remove high spatial frequency aberrations which cannot be removed by moving the SM and PM segments. One of the NGST mirror configurations under study by the University of Arizona uses deformable PM segments, with over 330 actuators behind the surface of each segment. Since NGST will be out of contact with its ground station for a 16-hour period each day, it will increase the amount of mission time available for science if the optical control software is run on board the spacecraft. NGST mission planning has not yet determined if the optical control software will be run from the ground or on the spacecraft.

The REE effort studies ways of advancing the state-of-the-art for computing in space. One of these ways may be to use relatively inexpensive commercial microprocessors instead of space qualified microprocessors. Because of the long lead time to create a high reliability, space radiation resistant, space qualified processor, such items are based on designs several years old. Thus they usually have much lower computing power than the latest commercially available processors in PCs and Mac computers. To compensate for the lower reliability and occasional radiation induced error, REE is developing fault tolerant operating system and application software that automatically checks for errors and recalculates results where errors have been found.

Applications

Phase Retrieval by Misell Algorithm

The Misell algorithm [3] is an iterative method by which the phase response of an optical system can by derived from a set of out of focus images of a known object. For the purposes of a space-based telescope, a perfect point source without the distortion of the atmosphere, is always available in the form of a single star. The phase response, or wavefront, of the telescope should ideally be constant across the system entrance pupil.

This algorithm was coded in "C" with Message Passing Interface (MPI) libraries used for parallel computation. The code was ported from the MasPar MPL code previously developed by Lyon [4].

The number of floating-point operations, required memory and disk space was calculated so that REE can anticipate the computer requirements for this application. This is shown is table 1. We have found that the Misell algorithm, acting on 4 simulated 512x512 point images at +/- 1, +/- 2 waves out of focus, converges in 40 iterations.

Phase Unwrapping

The phase map produced by the Misell algorithm is the phasor of the complex field at pupil. As the range of an arctangent function, values of phase are restricted to the interval $(-\pi, \pi)$ or +/- one-half wave even though the true wavefront may contain values outside this range. In order to operate the active optical control system, the full range of phase values must be recovered. This recovery process is referred to as phase unwrapping.

After comparing several algorithms for phase unwrapping the best performance was found to using the directed acyclic graph algorithm supplied by John Dorband of NASA. At its lowest level the algorithm compares each pixel of the phase map to its nearest neighbors. If a neighbor's phase differs by more than $\pi/2$, then the pixel is flagged for unwrapping. The directed acyclic graph forms an optimal path for the unwrapping.

Dorband's code was elaborated on in two ways

1. The average phase is set to zero. An arbitrary constant phase adds no information and does not affect optical performance, but should be removed so that the active optical system can retain maximum mechanical range.

2. The phase difference between different segments of the PM is set to the lowest possible value, modulo 2π. Since 2π can be arbitrarily added or subtracted from any phase, it is best to choose those values that minimize the difference between segments. The DM is not segmented and must compensate for phase across all the segments. Minimizing the inter-segment phase difference minimizes the amount the DM must be deformed.

Actuator Fitting by least-squared wavefront error

Once the unwrapped wavefront has been determined the telescope actuators must be moved in a fashion to minimize some measure of the optical performance error. Examples of these measures include encircled energy of the point spread function, spatial frequency weighted wavefront and least-squared wavefront. Most of the effort on this program has been on least-squared wavefront error. We assume that the active optical system is linear, that is, a single matrix, **R**, relates the actuator position vector, **a**, and initial wavefront, $\mathbf{w}_0$, to the resultant wavefront, **w**, as follows:

$$w = w_o + Ra$$

The actuator vector that minimizes the squared wavefront error, $||\mathbf{w}||^2$, is given by:

$$a = (R^T R)^{-1} R^T w_o$$

In practice, this equation is solved using the Cholesky method of solving for a symmetric matrix:

$$Ca = b$$

where:

$$C = R^T R$$

$$b = R^T w$$

Once **a** is found, it must be checked against the physical limits of actuator travel. In a functioning telescope, after the actuators are commanded to their new positions another set of point images must be acquired and the Misell algorithm run again to see if the wavefront has indeed improved to the desired quality. It is possible that the control loop may have to be run for several iterations, as there are tolerances in the system's calibration, noise in each detected image and finite numerical precision in the algorithm. We are studying the effects of these errors on the control loop. For 512 x 512 pixel wavefronts and 375 actuators, the matrix **R** requires 750 Mbytes to store in double precision. Our code currently generates a transcendental function with just a few stored parameters to define **R**. If it is expected that **R** is time-invariant, then $\mathbf{R}^T\mathbf{R}$ can be stored, using only about 274 Kbytes (single-precision) and avoiding the computationally expensive matrix-matrix multiply.

An implementation of actuator fitting for mission use must take into account several factors related to computing resources, such as:

1. Number of floating-point operations (FLOPs). Table 1 shows the approximate number of FLOPs required for the Misell algorithm and for actuator fitting. N is the number of pixels on one side of the phase map. The total number of phase map pixels is N^2. The phase unwrapping requires a small number of FLOPs compared to the other two algorithms.

2. Disk space required. Storing **R** explicitly requires 750 Mbytes. **R** must be stored if its values are updated by some calibration procedure. The columns devoted to DM actuators are rather sparse, containing significant values on only about 2 percent of the elements of the column. Work remains to be done to exploit the sparse nature of these columns to speed the algorithm and reduce the required disk space.

3. Memory required. Again **R** is by far the largest user of memory for the cases in which it must be handled explicitly.

Table 1: Floating-point Operation

Misell Algorithm

N=256, 40 iterations: 524 $\times 10^6$ FLOP

N=512, 40 iterations: 2.3 $\times 10^9$ FLOP

N=1024, 40 iterations: 10.1 $\times 10^9$ FLOP

Actuator Fitting

Assumptions: # of actuators = M = 375 for stiff PM system. M = 2200 for U of Arizona active PM.

Matrix-vector multiply: $b = R^T w$. $2N^2M$ FLOPs

N=512, M=375. #FLOP = 197×10^6

N=512, M=2200. #FLOP = 1.15×10^9

Cholesky solve: $M^3/3 + 4M^2 + 8M$ FLOPs

M=375. #FLOP = 18×10^6

M=2200. #FLOP = 3.6×10^9

Transcendental R generation (if necessary): $\exp(-\alpha r)\sin(\alpha r + \pi/4)$.

N^2M points x 65 FLOP/point for DM only

N=512, M=349. #FLOP = 5.9×10^9

N=512, M=2200. #FLOP = 37.4×10^9

R^TR multiplication (if necessary): N^2M^2 FLOPs

N=512, M=375. #FLOP = 3.7×10^{10}

N=512, M=2200. #FLOP = 12.6×10^{12}

References

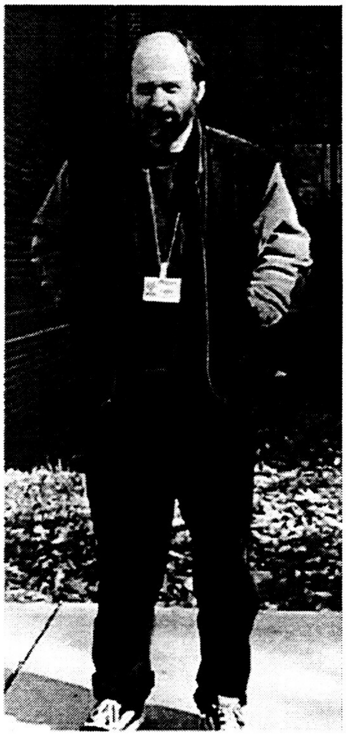
Stockman, H. S. ed. (1997). *The Next Generation Space Telescope*. Association of Universities for Research in Astronomy, Inc.

Lyon, R. G., Hollis, J. M., Dorband, J. E., and Murphy, T. P. (1998). Extrapolating HST Lessons to NGST. *Optics & Photonics News*.

Misell, D. L. (1973). A method for the solution of the phase problem in electron microscopy. *J. Phys. D*, 6, L6-L9.

Lyon, R. G., Dorband, J. E, and Hollis, J. M. (1997). Hubble Space Telescope Faint Object Camera Calculated Point Spread Functions. *Applied Optics, 36*(8).

Comparison of cloud data derived by TRIANA and numerical model simulation



Miodrag Rancic
University of Maryland Baltimore County
Department of Computer Science and Electrical Engineering
(mrancic@ciga.gsfc.nasa.gov)

In August last year, following a suggestion of Dr. Milt Halem (NASSA/GSFC), I began development of a high resolution numerical framework for assimilation of cloud data derived from the satellite TRIANA, that NASA plans to send to the point L1 in the year 2000. Originally, the collaborators on this project, include Dr. Halem, Dr. Fedor Mesinger (NCEP/NOAA) and Dr. Jules Kouatchou (Morgan State University).

This project has grown in scope, as will be elaborated bellow, and new participants have joined the team. Among them are Jim Geiger (GSFC) are Dr. Peter Norris (CESDIS). Dr. Sushel Unninayar has also given substantial input to the project through numerous constructive discussions.

As a part of collaboration with NOAA, for the major numerical tool we choose a massively parallel version of the NCEP regional Eta model, which is generally considered as a state-of-the-art in the regional weather forecasting, and has been particularly successful in precipitation forecast.

Project

We believe that in order to take full advantage of a dense mesh of cloud data that will be arriving from the TRIANA satellite (or, alternatively, that are presently arriving from TRMM), horizontal resolution of the numerical model has to be as high as approximately 10 km, and the model domain has to cover a whole path of the satellite around the globe. This, at least for the moment, appears to be practically unattainable.

Therefore, we formulated the following strategy, consisting of three components, or, hierarchy of models that are based on modifications of the NCEP Eta model.

1. A 10 km model, defined on the movable domain that will be constantly covering the sunlight side of the globe. The domain of this model should be tentatively spreading from -70 do +70 deg of latitude, and about 150 deg in the zonal direction. Presenting a movable domain on the massively parallel processors is the key issue in this set up, and we formulated a scheme which allows that locality of data with respect to the processors is maintained.

2. A 30 km model, defined over 70 deg belt around the globe, that should continuously provide boundary and initial conditions for the model run on the movable domain.

3. A 50 km full global model, that should generate a consistent set of boundary conditions for the belt model.

For preprocessing, that is, preparation of the initial fields, in the first stage we decided to use interpolations from the NCEP global analysis. The NCEP is developing an advanced 3D VAR method for initialization of the Eta model, but this code is still not mature enough to be applied for the global extensions required within our project.

The computation should be organized in such a way that all preprocessing and postprocessing is done on a workstation, but the model itself should be run on the Cray T3E.

Accomplishments

1. A portion of the preprocessing code for preparation and treatment of boundary conditions for the movable domain is finished. The details of the scheme for moving the domain on the massively parallel computer has also been worked out. However, we realized that the actual code development and its testing critically depend on preprocessing and the results of the belt model. Therefore, in this stage of development, we gave priority to this component of the project.

2. Though it was not the primary objective, we very soon became aware that the belt version of the Eta model by itself represents an important contribution, especially as a tool for studying the effects of increased spatial resolution on global scales. We successfully formulated and tested the belt Eta model in integrations up to 72 hours, on average horizontal resolution of 28 km. The preliminary results with the Eta-belt, not only clearly confirmed potential advantages of the belt-model concept in the long-term integrations, but also warned about limitations of the limited-area models. These results will be presented at the 13th Conference on Numerical Weather Prediction, organized American Meteorological Society, that will be held in Denver, CO, in September this year (Rancic et al. 1999).

Based on these results, we are planing to demonstrate general applicability of the Eta-belt model in:

- climate research, as a tool for down-scaling the results of the climate models;
- high-resolution simulation of volcano ash spreading;
- studies on global energy and water cycle.

An example of the Eta-belt forecast accomplished under this project, can be seen on page 78.

3. A new approach for treatment of the polar boundary problem is formulated. It consists of a 'wrapping over' technique for treatment of the polar points and formulation of a descending hierarchy of the overlapping grids near the poles, as a replacement of the standard polar filtering. Preliminary testing is in progress.

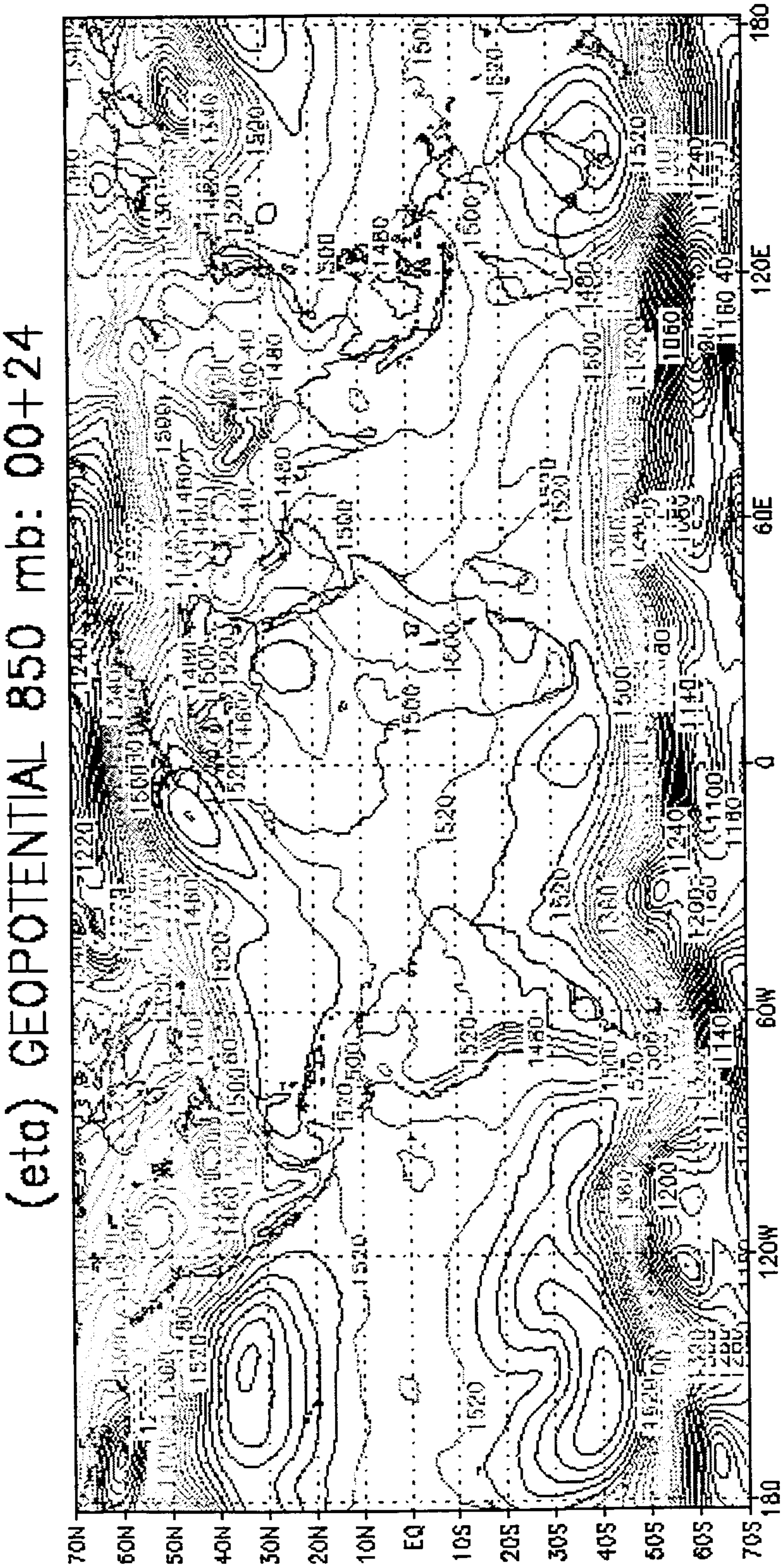
Current Developments

At the moment, we are working in the following directions:

- Jim Geiger is preparing a one month integration with the belt model;
- Peter Norris is preparing output of cloud data from the model;
- Jules Kouatchou is working on final details of preprocessing for the movable domain;
- Miodrag Rancic is preparing first tests with the global model, and is preparing an extension of the Eta-belt for simulation of volcano ashes in collaboration with Dr. Igor Eberstein.

References

Rancic, M., Geiger, J., Kouatchou J. and Mesinger, F. (1999). A belt version of the Eta model: Preliminary results. *Preprint Volume of the 13th Conference on Numerical Weather Prediction*, Denver, CO. 3.6.


(eta) GEOPOTENTIAL 850 mb: 00+24
70N
60N
50N
40N
30N
20N
10N
EQ
10S
20S
30S
40S
50S
60S
70S
180
120W
60W
0
60E
120E
180

Speedup to Virtual Petaflops using Adaptive Potential Solvers and Integrators for Gravitational Systems

George Lake
University of Washington
Department of Astronomy
(lake@astro.washington.edu)

Over the last few years, we've taken a code designed to simulate the origin of large scale structure and adapted it to follow the formation of the Solar System. Over the last decade, a Japanese Team has developed a special purpose computer that hardwires a simpler algorithm, but executes it at a Teraflop. Currently, our code simulates the system faster on a PC than their Teraflop system. We plan further advances to achieve a Virtual Petaflop.

Planet formation theories are modern versions of Kant's Nebular Hypothesis divided into stages where dust grains become kilometer-sized bodies by non-gravitational interactions and these planetesimals agglomerate into the present-day planets owing to gravitationally driven pairwise accretion (see Lissauer 1993 for a view of this fundamental picture that dates back only as far as Safronov 1969). However, models of planetesimal evolution have been forced to rely on analytical approximations, statistical techniques, or direct *N*-body methods with comparatively few particles and severe spatial restrictions. Comprehensive direct simulation must evolve a prohibitive number of bodies ($N \sim 10^6$-10^7) for an equally daunting time, ~ 10^6-10^7 orbits. We propose to perform such simulations, but let's first consider the scientific issues that motivate such effort.

We want nothing less than to accurately evolve a protoplanetesimal system into a planetary system that is qualitatively similar to our own Solar System. There should be a handful of widely-separated terrestrial planets with a relatively small range of masses all moving in the same sense in roughly circular orbits and in more or less the same plane. Of course, we won't find an exact match, so we will rely on our largest simulations to provide realistic distributions of masses, eccentricities, and inclinations for smaller suites of simulations (the minimum number of particles suitable for these studies will be based on the results of the larger simulations). Such simulations will provide insight and constraints regarding planetary systems around other stars, particularly clarifying the role of giant planets in the formation of terrestrial planets that are hospitable to life. In achieving this broad overarching goal, we will address several fundamental questions regarding the formation of planetary systems:

What are the planet formation timescales? The timescales will be sensitive to the initial mass distribution and the nature of the growth processes (i.e., whether there was a period of runaway growth). However, there are important observational constraints. For example, optically thick disks around pre-main-sequence stars become optically thin in, ~ 1-10 Myr (Strom et al. 1993), setting a limit to the timescale for the planetesimals to become large enough for "grinding" collisions to return dust to the disk. Subsequent evolution from these protoplanets to planets in the inner Solar System may take significantly longer (~ 10^8 yr). The transition from rapid growth to long-term interactions has been treated only qualitatively so far. Fundamental timescale constraints will reappear throughout our discussion.

What was the primordial surface density? By smearing out the known mass in the Solar System and allowing for depletion of volatiles, we can use the current state to guess the initial surface mass density distribution $\Sigma(r)$ in the inner nebula. We can then see if a higher density is required.

What controls "runaway" growth? While it is now generally accepted that a few bodies do detach from the general planetesimal mass distribution with accelerated growth rates after a certain amount of time and under certain conditions (e.g. the form of the mass and velocity distribution is important), some of the details remain uncertain. This is because direct simulations have to date been too coarse to do more than scratch the surface of the problem. Our numerical simulations will have sufficient dynamic range and time coverage to quantify directly the conditions under which runaway growth both begins and ceases to become effective.

Was there strong radial mixing? Past simulations suggest radial mixing during protoplanet accumulation sufficient to blur chemical gradients—at odds with the dependence of asteroid spectral type on semimajor axis. Our simulations will provide a detailed picture of radial mixing by merely comparing initial and final orbital radii.

What determines planetary spin? Six of our planets have spin vectors aligned with the common orbital vector, while the remaining three (Venus, Uranus, and Pluto) are retrograde.

Why is the asteroid belt so sparse? There is only 3 x 10^{24} g of material between 2.1 and 4 AU. The size distribution is collisionally evolved and the characteristic relative velocities (~ 5 km s^{-1}) are larger than the escape velocity of even the largest asteroid, Ceres. The blame for thwarting accretion and carving "the gaps" is nearly always attached to Jupiter. The first requires the rapid formation of Jupiter (see above). The latter may face problems with the extent of mass depletion compared to the narrow width of the resonances, unless Jupiter's semimajor axis migrated during its evolution so the narrow resonance zones swept through the belt and ejected sufficient material.

Why are the planetary orbits so cold? N-body simulations of the late stage of terrestrial planet formation typically produce planets whose mean eccentricity and inclination are significantly larger than Earth or Venus. Gravitational coupling to smaller bodies may be necessary to damp down the eccentricity and inclination of the smaller bodies.

Why are many planets in or near resonances? One of the major dynamical features of the solar system is the Great Inequility: Jupiter and Saturn are very near a 5:2 mean motion resonance. Neptune and Pluto are in a 3:2 mean motion resonance. Locking orbits into resonances generally requires dissipation, and it has been suggested that scattering of planetesimals by the giant planets has caused the resonance locking of Neptune and Pluto (Malhotra, 1993). A quantitative study of this mechanism requires a direct simulation of the disk-planet interactions.

In addition to these issues there are fundamental questions regarding underlying physical processes that our investigation will be able to address. Identifying the factors that control runaway growth (see above) is one example. Another is the question of whether dynamical friction has been confused in the past with collisionally-damped equipartition of energy. Still another is whether oligarchic growth of planetary embryos is properly described as being driven by migration or by chaos. Finally, our code can easily be generalized for problems in planetesimal dynamics that are distinct from strict solar system formation, such as the formation of the Jovian moons, or the dynamics in Saturn's rings.

Our planetesimal code starts with the UW cosmological *N*-body code *pkdgrav*. The code was developed by an interdisciplinary collaboration of astrophysicists, applied mathematicians, and computer scientists. It is used by several groups for a wide variety of problems in the formation and evolution of galaxies and large scale structure. Parts of it are also used to follow the molecular dynamics of water and the folding/denaturing of proteins. The code employs spatial and temporal adaptivity to handle up to 10^8 particles with large dynamic range (density contrasts of 10^5, dynamical times varying by 10^3). The parallel implementation efficiently divides the work and memory requirements to achieve nearly linear speedup on MPPs.

Spatial adaptivity is achieved with a tree-code. The complete parallel code achieves a sustained performance on a 512-node Cray T3E that is ~ 100x that of a Cray C90.

This code has required several modifications to enable it to handle planetesimal integration. The first is a trivial conversion to double precision as single precision was sufficient for cosmology: planetesimals are R ~10-100 km in size and range over distances of ~ 1.5×10^9 km; timesteps can be as small as hours or days in simulations that last $\geq 10^6$ yr. This larger dynamic range necessitated the change to double precision, but also leads to increased gains from spatially/temporally adaptive algorithms.

The other minor modification was the addition of external potentials to include the central force of the Sun and ultimately gas drag. (Giant planets are represented as individual particles to capture the back-reaction of the planetesimal disk.) Since the planetesimals are in a disk geometry, the number of particles and cells that are required to reach high precision is nearly an order of magnitude less than in the cosmological simulations.

The more difficult additions to the code have been collisions and an integrator optimized for the strong central force of the Sun. Collisions need to be detected explicitly (the cosmological code uses softened forces that render "collisions" meaningless). Collision outcomes will be determined based primarily on the energy of relative impact (as well as other factors such as the impact parameter), the lowest energies generally leading to mergers and the highest energies leading to fragmentation.

We use the *pkdgrav* data structure to handle individual particle timesteps, but the integration scheme draws heavily on our group's work on evolving the Solar System for its lifetime. The integrations are always done with symplectic integrators that give "exact solutions to approximate Hamiltonians", so they provide strict bounds on variations of conserved quantities over many dynamical times and insure that no spurious dissipation can cause artificial orbital migration. Although these are low-order integrators, their advantages outweigh the need to use timesteps that are smaller than in a higher order method. Symplectic methods are made possible by dividing the Hamiltonian, which also enables us to invent powerful new integrators to speed the calculation.

The advances offered by importing the algorithms and hardware used for computational cosmology are remarkable. We redid calculations done by the Japanese Harp-2 group. A PC costing $1,500 outperformed their Teraflop machine by an order of magnitude.

We are now working on algorithmic gains for an addition factor of 100 speedup. Our eventual goal is to make a PC as fast as the next generation In this way, we will achieve of goal of first principle simulations of the formation of the solar system.

HPCC/ESS

Adam Frank
University of Rochester
Department of Physics and Astronomy
(afrank@alethea.pas.rochester.edu)

Report

Due to responsibilities for the University of Rochester I have not billed the contract for any days in the first half of 1999. My first bill was sent in for June.

While I do not, therefore, have anything directly to report yet I would like to note an important achievement from last year's contract. An article I wrote for Astronomy Magazine based on the research of the Malagoli, Gardner and Gombosi groups was awarded the American Astronomical Society's Solar Physics Division award for popular writing by a scientist. This article called "Blowing in the Solar Wind" focused on HPCC efforts in the area of Coronal Mass Ejections and highlighted the need for high performance computing to solve this critical problem. The recognition by the AAS Solar Physics Division shows that outreach efforts for this contract are being recognized by the scientific community as well as the general public.

For the second half of 1999 I plan on writing four articles. Two for the popular press and two for internal NASA publications. I have already contacted Judy Colon about a story for Insights magazine. I am also working on proposals for a story on the "New Sun" for National Geographic which would, again, focus on HPCC solar physics teams as well as story on micro-gravity fluid dynamics (Carey's group) for Popular Science. I have already worked quite hard on the latter story for Discover but they have changed management and so far they have rejected the idea.

OUTREACH AND EDUCATION TEAM

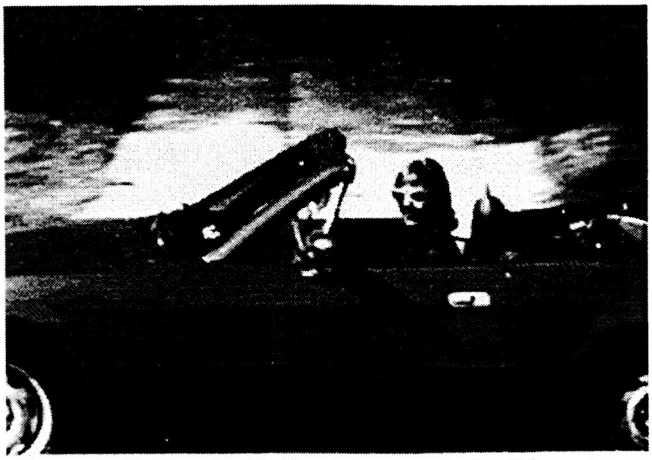
Susan Hoban, Acting Associate Director
(shoban@pop900.gsfc.nasa.gov)

Les H. Meredith, Senior Scientist
(les@usra.edu)

Sushel Unninayar
University of Maryland Baltimore County
Department of Computer Science and Electrical Engineering
(sushel@cesdis.usra.edu)

CESDIS Seminars

Digital Libraries



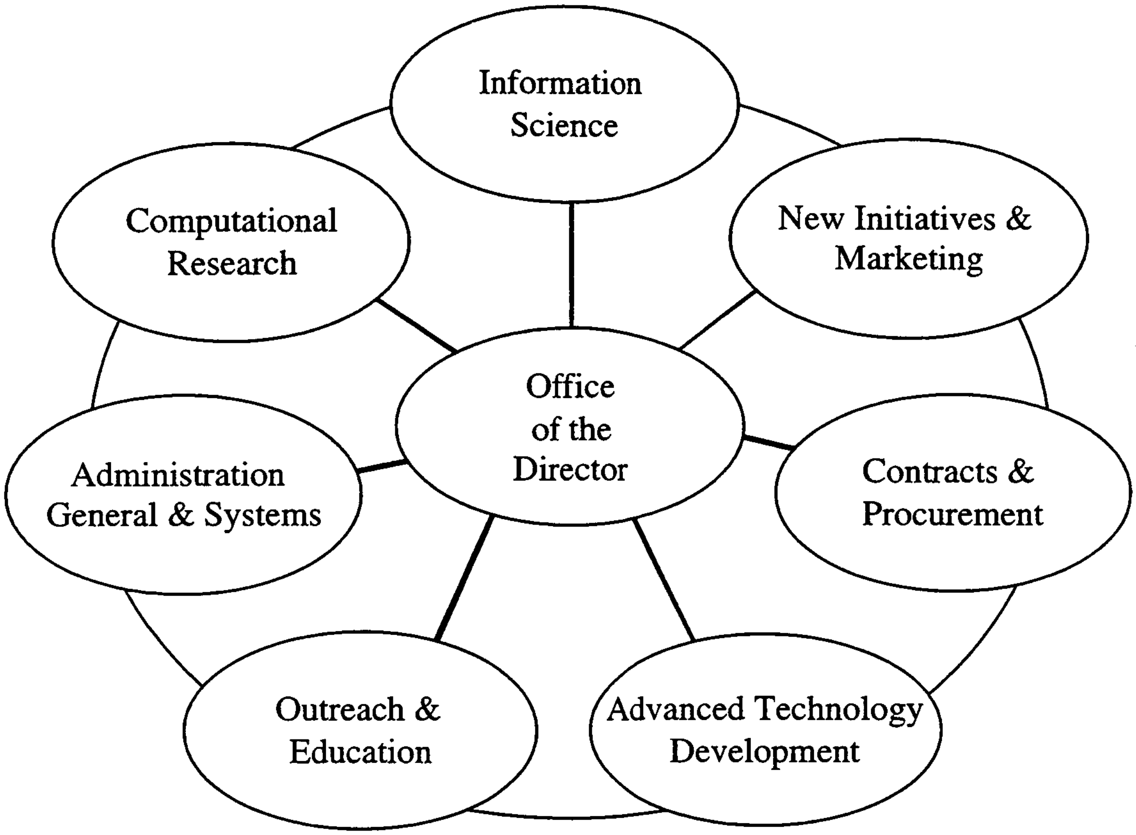
Susan Hoban, Acting Associate Director
University of Maryland Baltimore County
(shoban@pop900.gsfc.nasa.gov)

Office of the Director

In July 1998, Dr. Susan Hoban was appointed Acting Associate Director of CESDIS. In this position, Dr. Hoban supports the Director in the daily operations of the organization, coordination of the CESDIS Seminar Series, and represents CESDIS at USRA and NASA meetings, as well as at conferences and other outside activities.

CESDIS Realignment

In March 1999, in an attempt to align CESDIS with strategic changes taking place at Goddard and within NASA and the community, the Director and the Acting Associate Director implemented several organizational changes at CESDIS. The previous branch structure was transformed into a team-based structure, with a de-emphasis on organizational hierarchy and an emphasis on cross-disciplinary teams.



CESDIS Organizational Chart

Greater interaction among CESDIS scientists and between CESDIS and Goddard scientists, and increased collaborations with scientists outside of the Goddard community are among the goals that the realignment is meant to support. From the brainstorming sessions with the Teams, several ideas for approaching these goals have surfaced. One new implementation resulting from the realignment is *CESDIS Fruits & Java*, a gathering which typically preceded a CESDIS Seminar, to which a broad cross section of the Goddard Information Science and Technology community was invited. This activity, albeit small, was seen as a successful first step in facilitating communication among groups which previously were fairly unconnected.

Special Events

Among the notable special events hosted by CESDIS during FY99 were the Director's Special Seminar "Environmental Applications of Remote Sensing: Fire Detection and Modeling" and the Workshop on the "Roles of Computer Simulation." The Special Seminar was attended by NASA scientists, as well as international scientists from universities and private industry. The Simulation Workshop was held in recognition of the tenth anniversary of CESDIS, and was attended by scientists from NASA, U.S. universities and private industry.

Learning Technologies

(formerly Digital Library Technology)

The primary responsibility under this task is support of Dr. Nand Lal (NASA GSFC Code 933) in activities pertaining to the NASA HPCC Learning Technologies Project. Responsibilities include participation in bi-weekly teleconferences with LTP management at NASA/ARC, bi-monthly video-conferences with the LTP Inter-center Working Group, and trips to NASA/ARC and elsewhere for the Independent Annual review, LTP Annual Conference, LTP Advisory Panel meetings, LTP Retreats and other related meetings and conferences. The LTP effort at Goddard is focusing on the release of the LEARNERS cooperative agreement notice.

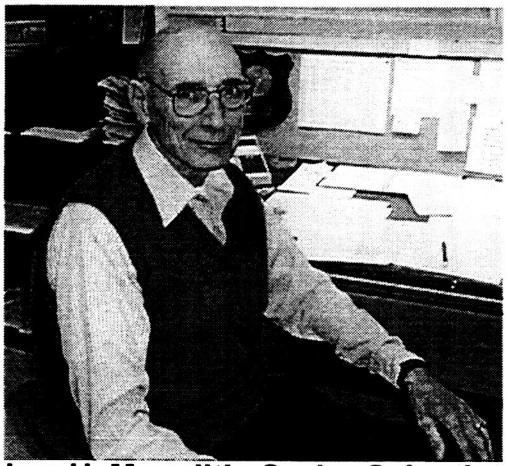
LEARNERS Cooperative Agreement Notice

In FY99, the primary activity has been development, release and review of the LEARNERS Cooperative Agreement Notice (CAN), a call for proposals for the development of educational technologies using NASA data as content. The LEARNERS CAN was written during FY99, and released on April 20, 1999. The solicitation was managed electronically. Proposals were received in May 1999 and reviewed in June 1999. Dr. Hoban supports Dr. Lal in all aspects of this process.

Miscellaneous

- **Digital Earth**: participating in the formulation of a program plan for the NASA participation in the interagency Digital Earth program.
- **Information Technology for the 21st Century (IT^2)**: supporting W. Campbell (NASA/GSFC, Code 953, Head) in development of a formulation plan for Goddard's participation in the IT2 program. Participated in several briefing to the Goddard Friends of Information Science and the Goddard Management Council.
- **JPL/LaRC/GSFC Knowledge Management Proposal**: Supported J. Bennett (NASA/GSFC Code 933, Head) by preparing materials discussing digital library technologies as applicable to Knowledge Management for a proposal to the NASA Chief Information Officer. The proposal was selected for award. CESDIS will be collaborating on this project in the area of intelligent information retrieval.
- **ADL99 Local Chair**: Served as Local Arrangements Chair for the Advances in Digital Libraries ADL99, held in Baltimore, MD, May 19 - 21, 1999.
- **Goddard Information Science and Technology Team**: serving on this team, which is tasked to develop a strategic plan for Information Science and Technology at Goddard for the next 5 years.

EXECUTIVE SECRETARIAT TO THE DATA AND INFORMATION MANAGEMENT WORKING GROUP OF THE U.S. GLOBAL CHANGE RESEARCH PROGRAM



Les H. Meredith, Senior Scientist
(les@usra.edu)

The Data and Information Management Working Group (DIMWG) acts as the data management arm of the U.S. Global Change Research Program (USGCRP) and provides an informal mechanism for interagency coordination and cooperation. Working Group agencies are the Department of Commerce, the Department of Defense, the Department of Energy, the Department of the Interior, the Environmental Protection Agency, NASA, the National Science Foundation, and the U.S. Department of Agriculture. The Department of State and the National Academy of Sciences serve as liaison members. The Data and Information Management Working Group has six subgroups and more than 50 active participants. The DIMWG supports collaboration between computer and Earth scientists involved in database, data management, and data distribution research by facilitating access to global change-related data and information in useful forms.

This task was assigned to CESDIS through the Global Change Data Center (GCDC) in the NASA Goddard Earth Sciences Directorate (Code 900). It requires the provision of Executive Secretariat support to the Data and Information Management Working Group including the guidance and coordination necessary to ensure future accomplishments which can be endorsed by the National Academy of Sciences and which enhance the level of general cooperation and participation of the DIMWG agencies. Les Meredith and is responsible for providing the support required by this task.

Profile

Dr. Meredith holds Bachelors, Masters, and Ph.D. degrees from the State University of Iowa. He is a Fellow of the American Association for the Advancement of Science, a Fellow of the Royal Astronomical Society, and a member of the American Geophysical Union, the American Physical Society, Phi Beta Kappa, and Sigma Xi.

Dr. Meredith's contributions to space science span more than 40 years and include employment as Head of Rocket Sonde Branch and Meteor and Aurora Section of the Naval Research Laboratory and a variety of positions at NASA Goddard Space Flight Center including Space Science Division Chief, Deputy Director of Space and Earth Sciences, Assistant Director, Acting Director,

Director of Applications, and Associate Director. He spent a year as Liaison Scientist for Space Science in Europe with the Office of Naval Research in London, four years as the General Secretary of the American Geophysical Union, and more than five years as its Group Director for meetings and advocacy.

Dr. Meredith is the recipient of the NASA Exceptional Scientific Achievement Medal (1965), the NASA Outstanding Leadership Medal (1975), the Senior Executive Service Presidential Meritorious Award (1981), and the NASA Distinguished Service Medal (1987).

1. Sent out meeting announcements, distributed the agenda and background material that I formulated, and wrote and distributed the minutes for ten DMWG meetings. The DMWG is successful in being the longest operating working group of the SGCR.

2. Formulated a concept for near-term implementation of a major new Global Environmental Change Information Service, GECIS, whose general concept is included in the USGCRP's strategic plan as described in Our Changing Planet 2000 and has been discussed with OSTP and OMB. The DMWG has agreed that this concept would be a basis for initial DMWG and agency planning.

3. Initiated and wrote the privacy policy that has been incorporated into the DMWG's Global Change Data and Information System Web page, GCDIS.

4. Drafted the data management policies for the SGCR National Assessment Working Group. They were approved and contractually implemented in all their four sectors and twenty national regions.

5. Drafted the DMWG's plan for the next few years that was requested by OMB through the SGCR. This plan included background material, interagency coordination, response to advisory reports, goals, near-term objectives, and performance measures.

6. Suggested to the DMWG and subsequently organized a "Special" DMWG. This meeting was Special in that the attendance included not just the regular DMWG agency attendees but senior people from the agencies with Global Change related data management programs. The meeting was very successful as evidenced by the special invitees asking that similar Special DMWG's be held twice every year in the future.

7. Developed and distributed a summary of the DMWG's activities in 1998 that included not only the DMWG's background but historical and recent accomplishments and summaries of all its 1998 meetings.

8. Established the criteria for data set inclusion, worked with the DMWG agencies to get citations for their data sets, and did all the necessary formatting for publication of a document entitled "1998 - Newly Available Agency Data Sets that are Significantly Global Change Related." It was published and is on the Web.

9. Drafted of a letter that the DMWG sent to OMB with minor modifications relative to OMB's proposed inclusion in their data access policy, A-110, of FOI coverage for data produced with Federal awards. The letter opposes the FOI language but emphasizes the importance of such data being made available. As an example, it gives the new award language to require this availability that the DMWG drafted in 1997.

10. Initiated and produced the data management policy section for GCDIS. This section includes the data management policy related actions over the last eight years of the DMWG, National Academy of Sciences, international scientific groups, OMB, UN WMO, EU, WIPO, and Congress as well as a section on recent comments and opinions, which I maintain.

EXECUTIVE SECRETARIAT TO THE COMMITTEE ON ENVIRONMENTAL AND NATURAL RESOURCES (CENR) TASK FORCE ON OBSERVATIONS AND DATA

Sushel Unninayar
University of Maryland Baltimore County
Department of Computer Science and Electrical Engineering
(sushel@cesdis.usra.edu)

The function of the Secretariat is to act on behalf of the CENR Task Force as the primary CENR interface for international consultations on scientific planning and implementation of the Global Observing System and its related data management system. This includes coordination with the international efforts underway be the Global Terrestrial Observing System (GTOS), the Global Climate Observing System (GCOS), the Global Ocean Observing System (GOOS), the Committee on Earth Observation Satellites (CEOS), the World Climate Research Programme (WCRP), and the International Geosphere-Biosphere Programme (IGBP).

This task was assigned to CESDIS through the Global Change Data Center (GCDC) in the NASA Goddard Earth Sciences Directorate (Code 900). It requires the provision of all the necessary technical and administrative support to assist the CENR Executive Director in implementing the responsibilities of the Secretariat. This includes coordinating the activities of the Task Force and its working groups, planning and coordinating U.S. participation in the International Global Observing System in accordance with the strategy outlined in the OSTP concept paper on the GOS, coordinating relevant observations and data management budget justification and advocacy material among the CENR subcommittees for submission to the Task Force, and coordinating with the Task Force's Data Management Working Group to promote effective access data management systems for CENR relevant global, regional, state, and local environmental and natural resources data.

Sushel Unninayar is responsible for providing the support required by this task. He works with CESDIS through a subcontract with the University of Maryland Baltimore County.

Primary functions and activities involved: (1) Executive Secretariat for the US Global Change Research Program's (GCRP) Interagency Working Group on Observations and Monitoring; (2) Scientific Advisor on the US Delegation to the United Nations Committee on the Peaceful Uses of Outer Space (COPUOS); (3) Executive Secretariat to the GCRP Global Water Cycle Program (and NASA/ESE Global Water and Energy Cycle Program); (4) Secretariat to NASA/ESE for the transitioning of research observing systems to operational platforms; (5) Co-chairman and organizer/coordinator of the UNISPACE-III/NASA Symposium on Climate Variability and Global Change; (6) Interagency coordination regarding the Integrated Global Observing Strategy (IGOS); (7) Coordination with the National Academy of Sciences' National Research Council, and the programs and projects of the World Climate Research Programme (WCRP), the World Meteorological Organization (WMO); (8) Coordination with other Federal Government agencies.

The 1998/1999 period was particularly notable for the extensive interactions involving the White House Office of Management and Budget (OMB) and the Office of Science Technology and Policy (OSTP) regarding the development of the long-term scientific plan for the US Global Change Research Program (US-GCRP). NASA was the designated lead agency for the GCRP global observing and monitoring program in which NASA/ESE is predominantly engaged through the various implemented and planned satellite missions. High level policy guidance caused a reorganization of GCRP scientific program plans to include new initiatives directed at research, observations and modeling of the global carbon cycle, the global water cycle and the ecosystem impacts of global change. These were in addition to the continuing thrust on ozone and atmospheric chemistry issues, particularly the monitoring of ozone depletion for the monitoring of the efficacy of the regulation of ozone depleting substances as embodied in the Montreal Protocol. Another significant realignment of previous scientific priorities related to the integration of research, observations and modeling to address climate on all time scales rather than the earlier separation of climate themes into seasonal/interannual variability and long-term climate change. An interagency working group was formed which combined the tasks of the CENR Task Force on Observations and monitoring and the USGCRP efforts in the same area. A Long-term strategic plan was prepared. Concurrently, the implementation plans for "Our Changing Planet (OCP)-2000" was prepared and submitted to the GCRP and thence to OMB and OSTP.

Intense efforts continued towards the finalization of the United Nations Draft Report of UNISPACE-III, which was scheduled to be held in July 1999, and the UNISPACE-III Conference Declaration. UNISPACE-III, the third United Nations Conference on the Exploration and Peaceful Uses of Outer Space represented a major international effort to review activities that occurred over the past 17 years with a view to set the stage for the next millennium. NASA was designated the lead agency to head the US delegation to the UN Committee on the Peaceful Uses of Outer Space (COPUOS) under whose auspices UNISPACE-III was being organized. As the chief scientific advisor on Earth observations, Earth sciences and environmental issues on the US delegation to COPUOS and UNISPACE-III, I was involved in all phases of the preparation for the Conference. I also was the Co-Chairman and Coordinator/Rapporteur responsible for organizing a special NASA/UNISPACE-III Scientific Forum on Climate Variability and Global Change. The proceedings of the Forum was published at Goddard and distributed in June 99, in advance of the UNISPACE-III Conference.

The Global Water Cycle was designated as one of the new initiatives of the USGCRP following guidance from OMB and OSTP. NASA was chosen the lead agency for this interagency program expanding along the lines and scope of NASA's global water and energy cycles program within the Earth Science Enterprise. The primary thrust of this new program is directed at the improved understanding, monitoring, modeling, and predicting the numerous aspects of the global water cycle involving interactions between the atmosphere, land surface and vegetation, and the oceans. Particular emphasis will be placed on the monitoring and prediction of water resources and water availability. Both space-based and in-situ observing platforms are involved with several new satellites such as TRMM contributing substantially to the quantification of the hydrological cycle. Initial efforts have been completed to form a scientific advisory working group to develop plans for the program, with initial planning beginning in the FY1999 (June/July) and a more complete program implementation strategy in FY2000 and beyond. The program is identified as a priority line item in OCP-2000 and will be followed up with additional resources in FY-2001 and beyond. The appointment of the scientific committee chairman and members and a first planning meeting is scheduled for late summer/early fall 99.

Also during 1998/99 a draft scientific implementation plan for NASA's Earth Science Enterprise was developed following the new thematic research focus areas of the reorganized Global Change Research Program, namely: Ozone and atmospheric chemistry; Climate variability and prediction across all time scales; Biology and biogeochemistry of ecosystems and the global carbon cycle; The global water cycle; and Solid Earth sciences (which includes natural hazards). The draft plan will be revised in early Fall 99 following internal review and vetting by OMB. Moreover, responding to an exchange of letters between Niel Lane (Science Advisor to the President) and Dan Goldin and Jim Baker, the administrators of NASA and NOAA respectively, a fast track white paper was drafted with input from agency representatives on the subject of transitioning research to operational systems for the consideration of OSTP and OMB. Deliberations on the subject are in progress.

Other activities included: Continued interaction with the National Academy of Sciences on the climate-infectious diseases study and participation in the planning phase of the WMO/WHO international conference on climate and health scheduled for Summer 2000; Interagency coordination regarding the further development of the Integrated Global Observing Strategy, led by the international Committee on Earth Observing Satellites (CEOS) for space-based components; Coordination with the WMO/WCRP and other organizations involved in climate research and modeling/prediction, and the Global Climate Observing System. Travel included participation at meetings in Tucson, Arizona (Climate-health conference planning), and Vienna, Austria (United Nations/COPUOS).

Seminar Announcement

Wednesday July 9, 1998
Building 28, Room E210, 10:00 a.m.
Hosted by Dr. Yelena Yesha

On the Optimal Split Tree Problem

Dr. S. Rao Kosaraju
Department of Computer Science
Johns Hopkins University

We study a tree construction problem motivated by applications to internet access. This Optimal Tree Problem is a generalization of the classic Huffman Coding Problem, for which a simple polynomial time optimal algorithm is known. We show that our problem is NP-complete and analyze a greedy heuristic to its solution. We show that the greedy algorithm guarantees 0(log n) approximation ratio. We also present several other performance bounds for this algorithm.

This work is done jointly with Teresa Przytycka and Ryan Borgstrom.

S. Rao Kosaraju has been a faculty member at Johns Hopkins University since 1969, and he currently holds the Edward J. Schaefer Chair. He serves on the editorial boards of several journals including SIAM Journal on Computing, for which he has been an editor for over 22 years. He is a fellow of ACM and IEEE. In October, he chaired the 1997 ACM Fellows Selection Committee.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Kosaraju, please contact Shelly Meyett at 301-286-8755.

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Thursday September 10, 1998
Building 28, Room W230F, 11:00 a.m.
Hosted by Dr. Yelena Yesha

An Overview of Quantum Computation: Concept and Intuition

Dr. Samuel J. Lomonaco Jr.
Department of Computer Science & Electrical Engineering
University of Maryland Baltimore County

This talk will give an overview of quantum computation in an intuitive and conceptual fashion. No prior knowledge of quantum mechanics will be assumed.

The talk will begin with an introduction to the strange world of the quantum. Such concepts as quantum superposition, Heisenberg's uncertainty principle, the "collapse" of the wave function, and quantum entanglement (i.e., EPR pairs) are introduced. This part of the talk will also be interlaced with an introduction to Dirac notation, Hilbert spaces, unitary transformations, quantum measurement.

Simple examples are then given to explain and illustrate:

1) Quantum teleportation
2) Shor's quantum factoring algorithm
3) Quantum error-correcting codes
4) Quantum cryptography

If time permits, the speaker will not be able to resist the temptation of discussing more advanced areas in quantum computation.

More information on some of the above topics can be found in the speaker's Lecturenotes Volumes at the URL: http://www.cs.umbc.edu/~lomonaco/lecturenotes

Dr. Lomonaco's research interests span a wide range of subjects from knot theory, algebraic & differential topology to algebraic coding theory, quantum computation, & symbolic computation.

Dr. Lomonaco is internationally known for his many contributions both in Mathematics and in Computer Science. In mathematics, Dr. Lomonaco provided a solution to problem 36 of R.H. Fox, a problem that resisted solution for over 15 years. In doing so, he created the hyperbolic section representation of four dimensional knots, and a homology theory for systems of groups connected by morphisms. He also demonstrated that Saunders Mac Lane's algebraic 3-type completely classifies a large class of four dimensional knots. Recently, Dr. Lomonaco has shown how knot theory can be applied to solve some outstanding problems in electrodynamics. He also serves as an associate editor of the Journal of Knot Theory.

In computer science, Dr. Lomonaco has used group representation theory to develop the theory of non-abelian error-correcting codes. He has developed a symbolic algorithm for factoring integers that reduces integer factoring to the task of solving boolean equations. For his many contributions to the development of the programming language Ada, Dr. Lomonaco received an award from the United States Under Secretary of Defense for Research and Engineering, Dr. Richard DeLauer. In quantum cryptography, he has shown how quantum information theory can be used to gain a better understanding of eavesdropping with quantum entanglement.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Kosaraju, please contact Shelly Meyett at 301-286-8755.

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

CESDIS

Friday October 16, 1998
Building 28, Room E210
11:00 pm - 12:00 pm

"A study of nocturnal marine stratocumulus development using Lagrangian (particle-based) large-eddy simulation"

Dr. Peter Norris
National Institute of Water and Atmospheric Research

A completely Lagrangian (particle-based) model has been developed for atmospheric large-eddy simulation. The Lagrangian method has advantages 1) in the simulation of parcel processes such as cloud microphysics; 2) in the absense of formal spatial organizations and constraints imposed by grid methods; and 3) in the implicit ease of parcel trajectory analysis. The Lagrangian LES solves the dynamics using the ``Smoothed Particle Hydrodynamics" technique. Other components are a TKE-based sub-parcel-scale turbulence closure coupled to a Monin-Obukhov treatment of surface layer fluxes; explicit parcel microphysics; and an emissivity parameterization of longwave radiative transfer.

A simulation of the nocturnal development of a stratocumulus-topped marine boundary layer is favorably compared against aircraft measurements from an ASTEX case study. The
enchroachment rate of the inversion is well predicted, as are moisture and buoyancy fluxes.

A study of cross-inversion entrainment shows that warm dry air from the inversion is incorporated into the cloud in the downdraft branches of large eddies and subsequently pinched off to form pockets of inversion air within the cloud body. These pockets slowly diffuse into the cloud as they move within it.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Norris, please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Friday November 13, 1998
Building 28, Room E210
11:00 am - 12:00 pm

Dr. Patrick Kinney
Columbia University

"Ozone and Epidemiological Studies"

Data Mining (or KDD) has gained prominence in the last 5 years as a technology with great promise. However, much of the work in data mining has concentrated on scale -- how to mine vast aggregates of data on machines with (relatively) limited computational power and memory. The data itself has always been relatively simple with easily available "equality" measures. Our interest is in mining in situations where the equality is not clearly defined, and we have to deal with the notion of similarity instead. In this talk, we will present this problem and show how techniques from Computational Intelligence (such as Neural Networks and Fuzzy Logic) may be appropriate to deals with it. We will anchor the talk around "Web Mining" as an application, but will also touch upon other applications such as mining from image/video databases.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Norris, please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

CESDIS

Wednesday November 18, 1998
Building 28, Room E210
11:00 am - 12:00 pm

"Practical Approach to Guiding Large, Living Software Projects"

by

Irene Qualters

Software Projects tend to grow in complexity and size, live a long life and undergo significant structural change. This trend still seems to catch us unprepared. Predictability of delivery, reliability, performance are often elusive; yet literature, experience, effort, talent abound. This talk will draw on some successes and failures observed with creating new products or transforming existing software in a context of time, resource, technical constraints. The emphasis will be on approaches and expectations which are suitable for sustained delivery of software.

For further information regarding directions, or access to NASA Goddard Space Flight Center, please contact Yolanda Smith at 301-286-4403.

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Friday November 20, 1998
Building 28, Room E210
11:00 am - 12:00 pm

"Rotation and Translation-invariant Image Representation"

Professor Eero P. Simoncelli
Center for Neural Science, and
Courant Institute of Mathematical Sciences
New York University

Orthogonal wavelet transforms have become a popular representation for multi-scale signal and image analysis. One of the major drawbacks of these representations is their lack of translation invariance: the content of wavelet subbands is unstable under translations of the input signal. Wavelet transforms are also unstable with respect to dilations of the input signal, and in two dimensions, rotations of the input signal. I'll discuss overcomplete image representations that avoid these difficulties. In particular, I'll derive a generalized class of rotation-invariant linear operators, show a variety of examples of such operators, and demonstrate the use of these operators for problems in image denoising, edge and junction analysis, and texture synthesis.

Eero Simoncelli received a Bachelor's degree in Physics, summa cum laude, from Harvard University, studied Mathematics on a Knox Fellowship at Cambridge University, and received a Master's and PhD in Electrical Engineering and Computer Science from MIT. From 1993 until 1996, he was an assistant professor of Computerand Information Science at the University of Pennsylvania. He is currently an assistant professor at New York University. Professor Simoncelli received an NSF Faculty Early Career Development (CAREER) grant in September '96, for research and teaching in "Visual Information Processing", and a Sloan Research Fellowship in February of 1998.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Professor Eero P. Simoncelli please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Monday November 23, 1998
Building 28, Room E210
1:30 pm - 2:30 pm

Hosted by
Dr. Yelena Yesha

Data Mining, Web Mining, and Computational Intelligence

Professor Anupam Joshi
Department of Computer Science & Electrical Engineering
University of Maryland, Baltimore County

Data Mining (or KDD) has gained prominence in the last 5 years as a technology with great promise. However, much of the work in data mining has concentrated on scale -- how to mine vast aggregates of data on machines with (relatively) limited computational power and memory. The data itself has always been relatively simple with easily available "equality" measures. Our interest is in mining in situations where the equality is not clearly defined, and we have to deal with the notion of similarity instead. In this talk, we will present this problem and show how techniques from Computational Intelligence (such as Neural Networks and Fuzzy Logic) may be appropriate to deals with it. We will anchor the talk around "Web Mining" as an application, but will also touch upon other applications such as mining from image/video databases.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Professor Joshi please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Friday December 4, 1998
Building 28, Room W230F
1:00 pm - 2:00 pm

"Data Mining in Very Large Dimensional Data Sets"

Dr. Vipin Kumar
University of Minnesota

Data sets with high dimensionality pose major challenges for conventional data mining algorithms. For example, traditional clustering algorithms such as K-means or AutoClass fail to produce good clusters in large dimensional data sets even when they are used along with well known dimensionality reduction techniques such as Principal Component Analysis. Similarly, traditional classification algorithms such as C4.5 perform poorly on large dimensional data sets.

This talk presents a novel method for clustering related data items in large high-dimensional data sets. Relations among data items are captured using a graph or a hypergraph, and efficient multi-level graph-based algorithms are used to find clusters of highly related items. We present results of experiments on several data sets including S\&P500 stock data for the period of 1994-1996, protein coding data, and document data sets from a variety of domains. These experiments demonstrate that our approach is applicable and effective in a wide range of domains, and outperforms techniques such as K-Means even when they are used in conjunction with dimensionality reduction methods such as Principal Component Analysis or Latent Semantic Indexing scheme.

This talk also presents a graph-based nearest neighbor classification scheme in which the importance of discriminating variables is learned using mutual information and weight adjustment techniques. Empirical evaluations on many real world documents demonstrate that this scheme outperforms state of the art classification algorithms such as C4.5, Ripper, Naive-Bayesian, and PEBLS.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Norris, please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

CESDIS

Wednesday December 9, 1998
Building 28, Room E210
1:30 pm - 2:30 pm

"Strategies for four-dimensional variational data assimilation using the FSU Global Spectral Model with its full physics adjoint"

I. Michael Navon
Florida State University

We conducted four-dimensional variational assimilation (4D-Var) experiments by using both a standard method and an incremental method. The adjoint of full physical parameterizations was used in the standard 4D-Var, while the adjoint of selected physical parameterizations was used in the incremental method. We examined influences of physical processes on 4D-Var by comparing the results of these experiments. As a whole, the inclusion of full physics into the adjoint model was detrimental to the minimization process, which primarily resulted from the boundary layer physics. The precipitation physics in the adjoint model tended to become beneficial after iteration 50. We confirmed that the assimilation analyses from the full physics adjoint model displays a shorter precipitation spin-up time. However, the benefit to precipitation spin-up did not result solely from the precipitation physics in the adjoint model, but from combining influences of a few physical processes.

A minimization algorithm was introduced, aimed at circumventing the detrimental impact and finally taking into account the positive effect of the physics in the adjoint model. This algorithm was based on the idea of truncated Newton minimization methods and the sequential cost function incremental method introduced by Courtier et al. (1994), consisting of an inner loop and an outer loop. The incremental method comprises the inner loop,while the outer loop consists of the standard 4D-Var using the full physics adjoint. The limited-memory quasi-Newton method (L-BFGS) was used for both inner and outer loops, while the information on the Hessian of the cost function was jointly updated at ever iteration in both the loops. In a two-cycle experiment, the quality of the assimilation analyses is fully better than that obtained from the standard 4D-Var or the incremental 4D-Var. The CPU time increased by 35% compared with the incemental 4D-Var without physics in the adjoint model, while the standard 4D-Var with full physics adjoint model increased by more than 100%.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Navon, please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Friday December 18, 1998
Building 28, Room E210
10:00 a.m.

hosted by: Miodrag Rancic

The Eta Model: Design, Performance, Some Conclusions, Future

Dr. Fedor Mesinger
National Center for Environmental Function

The design of the Eta Model is summarized. Features of the model which are unique or are considered particularly beneficial are emphasized. Of the numerical schemes, these are the step-mountain vertical coordinate, Arakawa-type horizontal advection, gravity-wave coupling scheme, energy conservation in transformations between the kinetic and the potential energy in space differencing, and the lateral boundary conditions scheme.

The performance of the model over the past somewhat more than a decade since its coming to life at the then National Meteorological Center (NMC) is reviewed. Various inferences can be made from comparisons against the performance of other NMC, now NCEP, models; as well as from Eta experiments aimed at identifying the impact of a specific model feature. These address the choice between an Arakawa-style against several alternative numerical approaches; the validity of the limited-area concept; the domain-size vs. resolution trade-off; and the impact of the eta coordinate.

Overall progress achieved and the performance of the current operational 32-km Eta are reviewed. Opportunities for improvement, work in progress and expected future trends are commented upon.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Mesinger, please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Friday January 8, 1999
Building 28, Room E210
10:30 am - 11:30 am

hosted by: Don Becker

"Beowulf and other Clusters"

Dr. Ron Minnich
David Sranoff Labs

A Virtual Single System Image (VISSI) environment for Beowulf and other clusters . At Sarnoff we have built Cyclone, a 160-node cluster consisting of 128 dual-pentium nodes, 16 533 Mhz. Alphas, and a collection of 16 nodes ranging from Pentium II/450 machines to old P90s. We began work on Cyclone in 1994 with just 16 P90s, and it has grown since then. Eighty of our nodes run Linux, and the other 80 run FreeBSD.

The Cyclone work extends our work in clustering that began in 1991 on SPARCstation machines. In that time we have been able to gain an understanding of what applications programmers need to get work done on clusters, as well as what types of systems work well and what types fail.

Programmers in general want their programming environment to look like a Single System Image (SSI). Many attempts have been made over the last quarter-century to make this model work, starting with Farber's DCS Ring in 1973. Unfortunately, in practice, SSI does not scale well in a single cluster, much less across clusters and in distributed systems. Our answer to this problem is to build tools to support Virtual SSI, or VISSI, systems.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Minnich, please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Tuesday March 2, 1999
Building 28, Room E210
11:00am - 12:00 pm

"Visualizing the Earth using TerraVision II"

Dr. Yvan G. Leclerc
SRI International
Artificial Intelligence Center

TerraVision II, and its associated suite of tools, allows users to create and visualize very large terrain datasets distrbuted over a network. These datasets combine terrain elevation data, aerial and satellite images, and various 3-D models. These are stored in a tiled, VRML-based, level-of-detail hierarchy. This distributed hierarchy, combined with an efficient caching and rendering mechanism, is what allows TerraVision II to view very large datasets at rates of 20 frames per second or higher, independent of network bandwidth.

In this talk, I will present TerraVision II's data organization and details of its internal processing. This will be followed by a demonstration using locally stored data and discussions about its possible future uses.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Leclerc, please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Tuesday March 9, 1999
Building 28, Room E210
12:30pm - 2:30pm

The Information Power Grid Project: *Research, Development, and Testbeds for High Performance, Widely Distributed, Collaborative, Computing and Information Systems Supporting Science and Engineering*

William E. Johnston, NASA Ames and Lawrence Berkeley National Laboratory
Dennis Gannon, NASA Ames and Indian a University
William J. Nitzberg, NASA Ames Research Center
William Feiereisen, NASA Ames Research Center

Modern science and engineering requires large-scale computation, high volume data management, and sharable instrument systems, all integrated with human collaboration, and all available in widely distributed environments. These requirements are driven by the need to: - do computer modeling of complex phenomenon; acquire, organize, analyze, visualize, and move around the world, massive amounts of diverse data; couple instrument systems to large-scale computing and data management systems for real-time analysis, steering, and remote control, and for direct comparison of experimental and computational simulations, and; to provide computer mediated human collaboration that is integrated with software systems that assist in the human creative process.

The science and engineering community also presents unique computing and information systems challenges in the diversity of the problems that they address, the diversity of resources that must be used, and the fact that as problems and approaches change - sometimes relatively quickly - the required resources change. These resources - computational systems, data repositories, instrument systems, and human collaborators - are diverse in form and function, are geographically dispersed, and are independently administered. In order to act in concert to solve the complex, multi-faceted, and frequently transient problems of scientific and engineering R&D, the relevant resources must be dynamically located, interconnected, and integrated into logical systems that are effectively built on-demand to address a single problem, with the resources be released when the system has completed its task. Providing this sort of a computing and information systems environment to support NASA's diverse research and development activities is the goal of the Information Power Grid project.

Our approach involves a combination of tactics: We are building on the work of, and actively collaborating with, the "grid" computer science community (see /5/); we are using commercial products in the subsystems where possible, and are developing missing components as necessary; and perhaps most importantly, we are integrating all of this work into a prototype production testbed in which real and complex application systems will be built and the effectiveness of the approach evaluated. (The initial applications will be drawn from work in NASA's HPCC/Computational Aerosciences and IT/Advanced Computing and Network Systems projects, and the Astrobiology and Earth Sciences programs, in order to ensure a fairly comprehensive range of requirements.) In this talk we will give an overview of the motivation, architecture, implementation, and current status of the Information Power Grid.

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Tuesday March 16, 1999
Building 28, Room E210
11:00am - 12:00 pm

"Automated Image Registration with Parameter Adjustment"

Mark Lucas
ImageLinks, Inc.

An introduction to ImageLinks/AGIS value added remote sensing processing will be given. Also, we will discuss Sensor Based Modeling and Autonomous Registration, by presenting a visual walkthrough of the algorithms describing ImageLinks sensor-based approach, adjustable parameters, error analysis, projective geometries, and feature correlation for autonomous registration of multiple sensors.

Finally, a description of open source efforts involving remote sensing, including BeoWulf support, will be given (see www.remotesensing.org).

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Mr. Lucas, please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

CESDIS

Friday April 2, 1999
Building 28, Room E210
11:00am - 12:00 pm

"Message Passing and Parallel File Systems for Beowulf Machines"

Robert B. Ross
Clemson University

While Beowulf computing continues to grow in popularity, many capabilities available on traditional commercial parallel computers are still unavailable for this platform. In addition, questions still remain with regards to the scalability of Beowulf and the optimal approaches to such simple tasks as message passing and data storage on the system. The Parallel Architecture Research Laboratory (PARL) at Clemson University is investigating a number of aspects of Beowulf computing both at the application and system software levels.

The presentation will first give an overview of the work in progress at PARL. Two specific projects will then be covered in detail. First the results of an evaluation of message passing packages available for Beowulf will be discussed. Finally the Parallel Virtual File System (PVFS) will be described.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Mr. Ross please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Mini-Workshop Announcement

Thursday April 8, 1999
Building 28, Room E210
1:00 pm - 2:30 pm

The Advanced Computing Technology Center IBM Watson Research Laboratory

Mr. John Levesque
Director of Advanced Computing Technology Center

The Advanced Computing Technology Center (ACTC) was established within IBM's T. J. Watson Research Lab to investigate requirements for developing, porting, and optimizing scientific and engineering applications to advanced high performance computers such as the IBM RS/6000 Scalable Processing Systems (SP). Mr. John Levesque, Director of the ACTC will be conducting a Mini-Workshop at GSFC for interested NASA researchers and contractor personnel. The purpose of the workshop is to present the goals and interest of the Advanced Computing Technology Center and describe the work to date in developing tools and libraries required to port and optimize application programs for sequential execution, shared memory parallelization, distributed memory parallelization, and combined shared/distributed parallelization. Inaddition, IBM is interested in learning about specific GSFC User interest for HPC applications. Mr. Levesque will also share the findings of a recent 3-day workshop held in March at the San Diego Supercomputer Center on this same subject.

Prior to joining IBM, John Levesque was a founding principle in companies such as Pacific Sierra Corporation and Advanced Parallel Research (APR) and is a noted developer of advanced products used extensively by high performance supercomputer users.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Mr. Levesque please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

CESDIS

Tuesday April 13, 1999
Building 28, Room E210
11:00 am - 12:00 pm

"Integrating Scientific Datasets and Digital Libraries"

Robert E. McGrath
National Center for Supercomputing Applications
University of Illinois, Urbana-Champaign

Our group at NCSA, and myself in particular, are in somewhat unique position. For the past six years, our group at NCSA has been deeply involved with both digital library technology and also the world of scientific computing and data. This paper presents some personal views about the current relationship of libraries and scientific data archives, and what I think should be done next.

We are already seeing the emergence of new, all digital scientific publishing, which is creating a convergence of missions between journal publishers, libraries, and data archives. Conventional libraries are still trying to discover their role in this emerging digital world. At the same time, the volume an diversity of scientific data on line is exploding, leading to increased efforts to integrate data from many sources. I believe that digital libraries can and should play a key role for scientists by providing a unified environment for information discovery and access.

We have built a unique prototype system, in which both text and data resources may be searched with a single query, and data of many types as well as text can be effectively retrieved. I think our work shows that we have the technical means to integrate science data into digital libraries. Our technology is completely general, it is being applied to several disciplines. Since we use Z39.50, we already access information from many disciplines, including medicine, engineering, and space science. This kind of environment not only allows but almost forces interdisciplinary research.

While the software is basically ready, there is much work to be done in the areas of standards and access. This is an intellectual and social problem, more than a technical one. We programmers can implement any reasonable standard for expressing queries and results. But these standards need to be developed by scientific communities, who typically are neither funded nor prepared for such efforts.

Accessing scientific data remains a significant challenge, for which we don't have the software yet. The development of good models for describing data will make it much easier to create tools and infrastructure. The foundations of this work are being laid, and if and when good standards emerge, software will rapidly follow.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Mr. McGrath please contact Yolanda Smith at (301) 286-4403.

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Wedensday April 28, 1999
Building 28, Room E210
3:30 pm - 5:00 pm

Japan Gigabit Satellite Project

Dr. Takashi Iida
Deputy Director , Communications Research Laboratory of Ministry of the Posts & Telecommunications

The Gigabit Satellite Project is a new Japanese experimental communications satelite project that has been funded through the R&D stage. Dr. Iida believes there may be an international cooperation element in this experimental project.

F or further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Iida Yolanda Smith at 301-286-4403.

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Tuesday May 11, 1999
Building 28, Room E210
11:00 am - 12:00 pm

Phase-Diversity Technology: Wavefront Sensing and Imaging

Rick Paxman
ERIM International

Phase diversity is a data-collection and processing technique used to jointly estimates wavefronts and fine-resolution imagery from aberrated focal-plane data. This technology has enjoyed great success in fine-resolution imaging through atmospheric turbulence. We review this success and discuss transitioning phase-diversity technology to space-telescope applications.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Paxman please contact Yolanda Smith at 301-286-4403.

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Monday May 17, 1999
Building 28, Room E210
11:00 am - 12:00 pm

Generation of Terahertz radiation: comparative analysis.

Jacob B. Khurgin

In the recent years the region of electromagnetic spectrum with frequencies between 1 and 10THz had generated quite an interest in the scientific community due to a number of interesting applications in the remote sensing, communications and non-destructive testing and imaging. This frequency region lies at the boundary between what is usually thought as electronic and optical domains, and, presently, there is no efficient source of the THz radiation available. In the absence of electronic (transistor) or quantum (laser) source the best results are currently obtained by mixing the radiation of two powerful laser sources of much higher frequencies (100 THz or more) and obtaining the difference frequency signal in the 1-10 THz range. Various schemes for that had been suggested and demonstrated.

In this talk, the latest experimental results will be reviewed, and then the fundamental limitations on the efficiency of THz difference frequency generation in various schemes will be considered. Among the materials for THz generation we describe the photoconductors, semiconductor surfaces, bulk insulating crystals and semiconductors, and finally semiconductor quantum wells and superlattices. We shall make connection between nonlinear optical methods of difference frequency generation and coherent oscillations in quantized structures (including Bloch oscillations). We shall also compare various methods of THz power extraction-impedance-matched antennae, waveguides or simple dipole radiation.

The main conclusion of this talk will be that for a given set of requirements: frequency, power, duty cycle etc. a different combination of materials and geometries can be optimal.

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Monday May 24, 1999
Building 28, Room E210
2:00 pm - 3:00 pm

David Brunnell
Cinebase

This talk discusses the challenges of effective media management solutions and how the Cinebase2 architecture, services, and applications address these challenges. Cinebase2 was shaped by several technology imperatives:

-the ever increasing complexity of platforms, formats, and tools, which drives a component architecture;
-the complexity of managing complex media assets, which drives the use of an object model;
-the need to effectively create, revise, and manage assets, which drives integration of workflow.

Cinebase2 provides a rich set of services and an extensible set of applications for building media management solutions. Cinebase2 services include Content Services, Media Format Services, Descriptor Database Services, and Workflow Services. Cinebase2 is a distributed architecture, designed so that configuration can be easily extended and so that the underlying hardware, not Cinebase, limits media size and speed of transfer.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Mr. Brunnell please contact Yolanda Smith at (301) 286-4403.

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

Tuesday May 25, 1999
Building 28, Room E210
11:00 am - 12:00 pm

Adapting Scientists' Investigation Tools for Inquiry Learners: A Case-Study of Visualization in Earth Systems Science

Daniel C. Edelson
Institute for the Learning Sciences
and School of Education & Social Policy
Northwestern University

Computing technologies offer tremendous potential for science education reform. Investigation tools and scientific resources can help to transform science learning from the passive absorption of knowledge that characterizes current practice, to the active construction of understanding through engagement in meaningful activities. In our research, we have been exploring the use of visualization and data analysis tools to support inquiry-based science learning. Through an iterative design process, this research has identified challenges to the implementation of technology-supported inquiry learning in real classrooms and led to the development of strategies to overcome them. In this talk, I will describe WorldWatcher, a geographic visualization environment we have developed for learners, and present the Alternative Worlds Project, a curriculum unit focusing on global climate. I will use these to illustrate the challenges of technology design, curriculum development, and teacher preparation that any design must overcome if it is to bring the promise of science education reform into classrooms.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Mr. Edelson please contact Yolanda Smith at 301-286-4403

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Seminar Announcement

CESDIS

Monday 14, June 1999
11:00 am - 12:00 pm
Building 28, Room E210

Charles L. Seitz, Ph.D., President & CEO of Myricom Inc.

Myrinet -- Scalable Cluster Interconnect

Dr. Seitz will offer a technical exposition of Myrinet, its technology roadmap, its applications, and its role in the evolution of high-performance clusters.

Charles (Chuck) Seitz earned S.B., S.M. and Ph.D. degrees in electrical engineering in the 1960s at M.I.T.

After a period in industry, Seitz joined the computer science faculty at Caltech, where his research and teaching activities were in the areas of microelectronic-chip design and concurrent computing. In Seitz's concurrent-computing research, principally under DARPA sponsorship, he and his students developed the first multicomputer, the Cosmic Cube; devised the key programming and packet-routing techniques for the second-generation multicomputers; and transferred these technologies to industry. Seitz was elected to the National Academy of Engineering in 1992 with the citation "for pioneering contributions to the design of asynchronous and concurrent computing systems."

In 1994, Seitz, his Caltech research team, and two researchers from another DARPA-sponsored research project at USC Information Sciences Institute founded Myricom, Inc., a company dedicated to making the high-performance interconnect used in multicomputers available as a commodity product. Myrinet, a gigabit-per-second packet-communication and switching technology, is a direct descendent of multicomputer message-passing networks, but without restrictions on link distance or network topology. Myrinet is now used at many hundreds of customer sites, including in many of the world's premier cluster-computing installations.

For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Seitz please contact Yolanda Smith at 301-286-4403.

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

ADVANCED TECHNOLOGY DEVELOPMENT TEAM

Phillip Merkey, Senior Staff Scientist
(merk@cesdis.gsfc.nasa.gov)

Donald Becker, Staff Scientist
(becker@cesdis.gsfc.nasa.gov)

Erik Hendriks, Technical Specialist
(hendriks@cesdis.gsfc.nasa.gov)

Neil R. Helm
George Washington University
(helm@seas.gwu.edu)

James Wang
George Washington University
(jwang@SEAS.GWU.EDU)

L. Michael Hayden
University of Maryland Baltimore County
Department of Physics
(hayden@umbc.edu)

Terrence Pratt, Senior Scientist
(pratt@cesdis.gsfc.nasa.gov)

Udaya A. Ranawake
University of Maryland Baltimore County
Department of Computer Science and Electrical Engineering
(udaya@neumann.gsfc.nasa.gov)

Fran Stetina
Fran Stetina and Associates
(stetina@gsti.com)

Beowulf Parallel Workstation

Phillip Merkey, Senior Staff Scientist (merk@cesdis.gsfc.nasa.gov)
Donald Becker, Staff Scientist (becker@cesdis.gsfc.nasa.gov)
Erik Hendriks, Technical Specialist (hendriks@cesdis.gsfc.nasa.gov)

Profiles

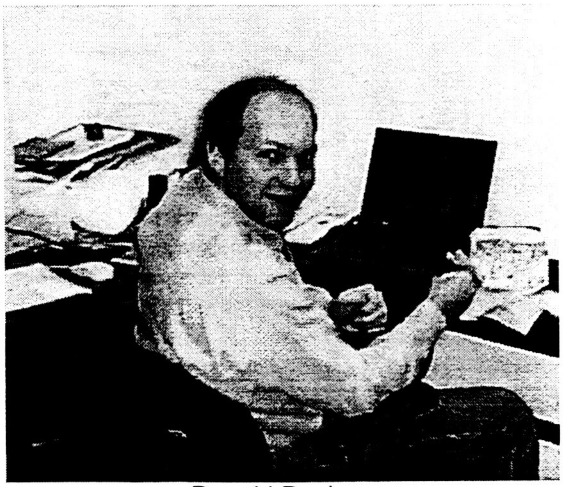
Phillip Merkey

Dr. Merkey holds a Bachelor of Science degree in mathematics from Michigan Technological University, and took a Ph.D. in mathematics in the area of algebraic coding theory from the University of Illinois (1986). He is a member of the AMS and SIAM.

Prior to joining CESDIS in 1994, Dr. Merkey was employed as a research staff member by the IDA Supercomputing Research Center in a classified working environment. His experience includes application of high performance computers to grand challenge problems, investigation of instruction level parallelism using the VLIW parallel computer, benchmarking experiments on the Multiflow Trace computer, algorithmic design for empirical solutions to problems in applied discrete mathematics, and innovative parallel implementations of advanced algorithms.

Dr. Merkey is the technical lead on the Beowulf Bulk Data Server project. He is responsible for the overall design and progress on the project. He also responsible for identifying and evaluating applications that will be suitable applications to demonstrate the machine capabilities and guide its development.

Dr. Merkey has also engaged in outside collaborations with the IDA Center for Computing Sciences, he has participated in Dr. Sterling's Petaflops workshops including studies of applications for the HTMT architecture and has served as an instructor at the University of Maryland Baltimore County were he is developing a course on Parallel and Distributed Computing and has taught the senior level Analysis of Algorithms course.



Donald Becker

Mr. Becker holds a Bachelor of Science degree from the Massachusetts Institute of Technology in electrical engineering and has completed graduate computer science courses at the University of

Maryland College Park. From 1987 to 1990 he was employed by Harris Corporation, Advanced Technology Department, Electronic Systems Sector as a senior engineer. He performed research and development work on the Concert multiprocessor, maintained and extended the Concert C compiler (based on PCC) and libraries, and wrote network software.

As a research staff member of the IDA Supercomputing Research Center from 1990 to 1994, Mr. Becker wrote a substantial proton of the low-level LINUX networking code, designed, implemented, and characterized an interfile optimization system for the GNU C compiler, implemented a peephole optimizer for a data-parallel compiler (DBC), and implemented several symbolic logic applications.

Since joining CESDIS in 1994, Mr. Becker has been the principal investigator for system software on the Beowulf Parallel Workstation project. He has established a world class reputation in the operating system community with his contributions in networking software. Mr. Becker continues to make CESDIS the center of the networking research community for Linux and Beowulf. He helped develop and has participated in several "How to build a Beowulf" tutorial sessions presented at leading conferences throughout the year. He is a co-author of "How to build a Beowulf", published by the MIT press.

Erik Hendriks

Mr. Hendriks received his Bachelor of Science degree in Computer Science from The Johns Hopkins University in 1996. During his graduate studies, he worked for the physics department at the Johns Hopkins University writing parallel programs.

Mr. Hendriks' primary responsibility is the development of system software for the Beowulf Project. Mr. Hendriks has continued to refined the installation procedure for Beowulf clusters, made significant contributions to the growing collection of Beowulf system software, has conducted an extensive evaluation of the candidate disks for the Bulk Data Server and has developed the software needed to run multiple disks at full aggregate speeds. Mr. Hendriks has improved his software that can access the hardware monitors on the motherboards used in the Bulk Data Server.

Mr. Hendriks has developed and released a software package called 'bproc'. The software addresses the ESS milestone for a global process id space. This approach uses ghost processes and PID masquerading to provide the functionality of a global process id space, but doesn't suffer from the scaling problems that plagued earlier attempts.

Addition to becoming an integral member of the Beowulf team, Mr. Hendriks has shown himself to be valuable member of CESDIS as well. On numerous occasions he took over responsibilities of the CESDIS system administrator and repaired or installed systems that enabled CESDIS to meet it diverse obligations.

Report

Beowulf Project continues to spread throughout the world and CESDIS continues to maintain a leadership role in the development of Beowulf Class Cluster Computing. The Website, http://www.beowulf.org and the associated mailing lists maintained by CESDIS continue to provide a focal point within the Beowulf community.

Don Becker has been a co-instructor at numerous tutorials on the construction of Beowulf Clusters. He has given numerous invited talks and tutorials across the country and internationally. Thomas Sterling, Don Becker, John Salmon and Daniel Savarese have co-authored the book

"How to Build a Beowulf" release by MIT Press this spring which captures this material in book form.

Don Becker received the Dr. Dobb award for excellence in software development.

Erik Hendriks has released and continues to develop one of the most exciting ideas within the Beowulf system software development effort. With bproc, processes run on the computation server nodes but appear, via the ghost processes, as if they are running on the head node. This is a big step towards making the nodes ``stateless computation servers" like they are in many MPPs.

Don Becker and Phil Merkey served on the program committee and as a session chairs for JPC4-4 (the 4th Joint PC Cluster Computer Conference) held in Pasadena, CA. This meeting brought together researchers from NASA, DOE, NIH, other agencies and a number of Universities.

CESDIS has continued to be the center of activity in network research, and with its web presence, has continued to be one of the major repositories for the Beowulf software and Beowulf technology.

Phil Merkey has refined the course on Parallel and Distributed computing based on the Beowulf technology. This course was again given in the fall semester at UMBC. After discussing the Parallel Computing from an academic point of view the students were given accounts on the Beowulf cluster called hrothgar. This "lab" component of the course provides hands on experience with parallel programming and debugging parallel programs and put the abstract analysis of parallel programs in a more tangible frame work.

Phil Merkey has taken on a leadership role in the development of the Round-3 of the HPCC/ESS program. Merkey will be replacing Terry Pratt, in his retirement, as the technical lead for evaluation and will lead the Beowulf Cluster Computing effort within the context of the Round-3.

The Beowulf Bulk Data Server has been upgraded to meet its design specifications. The cluster currently now has 128 Intel P6 processors running at 200 MHz and 8.2 GB of main memory and, 1.4 TB of disk storage. It is currently connected as one-half (the other half being John Dorband's, theHIVE) of a 256 processors Beowulf. The backbone for the systems is a set 72-port Foundry Switch connected by 4GBit/s Ethernet lines.

CESDIS has continued to work with researchers at Clemson University headed by Dr. Walter Ligon. This summer the team successfully ran the parallel filesystem PVFS on the Bulk Data Server while the theHIVE as the computation server.

Infrastructure Enhancements to the Global Legal Information Network (GLIN) and the Testbed for Satellite and Terrestrial Interoperability (TSTI)

Neil R. Helm
George Washington University
(helm@seas.gwu.edu)

Goals

1. To provide NASA and the Library of Congress with satellite communications planning and infrastructure for the Global Legal Information Network (GLIN). This involves procuring and assembly of the LOB ground terminals and providing satellite transponder access from a domestic satcom supplier.

2. To support the GIBN Trans-Pacific experiments with its implementation of the North American communications links, and to coordinate experiment activities with the Japanese team.

Work performed on the GLIN terminals and experiment

1. Successfully coordinated with PanAmSat, a domestic satcom provider, for free transponder time for the GLIN experiment.

2. Procured a communications equipment shelter for storage of our experimental equipment that needs to reside close to the LOB ground terminals that are located at the GSFC.

3. Assisted in the procurement of the two VSAT, Ku-band ground terminals for the LOB. Wrote the terminal specifications and participated in the team bid evaluation. The winning vendor was the Hughes Network Systems corporation.

4. Assisted in the delivery of the terminals. The inventory assessment determined that one major post mount was missing. Followed up successfully with vendor to deliver the missing part and thus completing the inventory.

5. Wrote a Terminal Assembly and Initial Test Plan for the ground terminals. Provided this test plan as a deliverable to the contract.

6. Conducted the orbital site analysis from the GSFC ACTS pad site to a number of prospective domestic communications satellites that may be used in our experiment. The results of the analysis was very positive with a clear elevation angle to the equatorial orbital arc.

7. Completed an Internet search of all the U.S. suppliers of domestic and international satellite communications services, and organized these data into a 40 page report and provided a short summation on its relevancy to our GLIN and TSTI networks. This 40 page report was a deliverable for the contract.

8. As part of this contract, I procured two Comsat CLA-2000/IP Link Accelerators for integration into the TSTI communications testbed at GSFC. The procurement consisted of a number of visits to Comsat Labs to review specifications and procurement procedures.

Work performed on the Trans-Pacific experiments

1. Conducted nearly weekly telecoms during 1999 with the US Trans-Pacific team.

2. Conducted numerous meetings in the US, Japan and Canada with the Japanese team members, coordinating both technical and applications parameters of the experiment.

3. Successful in getting AT&T to agree to be an experiment team member and allow us free access to its Salt Creek earth station in California. However, the costs of getting the data from NASA Ames to the Salt Creek terminal were later deemed to be prohibitive.

4. Discussions with Teleglobe of Canada and the Canarie research network were also successful, and it is now planned to use the Teleglobe Lake Cowichan ground terminal in Western Canada to carry the experiment data.

5. Worked with both Teleglobe, the Canadian Intelsat Signatory, and KDD, the Japan Signatory to Intelsat, to obtain free high data rate satellite communications for the Trans-Pacific experiment. The communications are via the hot standby cable restoration circuits.

6. Assisted the team in preparing end-to-end test goals and demonstration objectives.

James Wang
George Washington University
(jwang@SEAS.GWU.EDU)

Assisted Dr. Neil Helm in his work by assessing software products for web-based monitoring.

Holographic Storage Using Photorefractive Polymers

L. Michael Hayden
University of Maryland Baltimore County
Department of Physics
(hayden@umbc.edu)

During this period, we have been studying the feasibility of using photorefractive (PR) and photochromic (PC) polymers as holographic storage media for applications useful to the Earth and Space Data Computing Division (ESCDC) of the NASA/Goddard Space Flight Center and to NASA in general. During the past year, we successfully demonstrated the storage and retrieval of 50 multiplexed digital pages. We also studied several PR/PC polymer composite materials and measured many of their characteristics relevant to holographic storage. The results of that work were presented in a report to ESCDC on April 22, 1999.

The graduate student involved in that work, Shane Strutz, is currently supported by a NASA GSRP Fellowship which is scheduled for completion in June of 2000. During his Fellowship, he has studied the physics of the PR and PC effect in polymers, the material properties important to holographic storage, and built a demonstration holographic storage device.

Presentations and Publications

Hayden, L. Michael and Strutz, S. J. (1998). Co-located, permanent-photochromic and erasable-photorefractive holographic images, in *Xerographic Photoreceptors and Organic Photorefractive Materials IV*, S. D. Ducharme, J. W. Stasiak, (Eds.), *Proc. SPIE 3471*, 152.

Strutz, S. J. and Hayden, L. Michael (1998). Photorefractive polymer with both real-time optical processing and long term storage capability. *ACS Annual Mtg. Organic Thin Films for Photonic Applications*, Boston, MA.

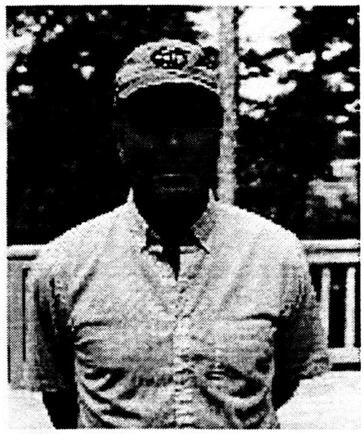
Strutz, S. J. and Hayden, L. Michael (1998). Photorefractive polymer with both real-time optical processing and long term storage capability. Post-deadline paper *OSA Annual Mtg.*, Baltimore, MD.

Strutz, Shane J. (1998). Photochromic and Photorefractive Polymers for Holographic Storage, invited talk to *Professor Demetri Psaltis' group at the California Institute of Technology.*

Hayden, L. Michael (1999). Hologram degradation in polymeric storage media (Invited). *International Workshop on Holographic Data Storage*, Nice, France.

Strutz, S. J. and Hayden, L. Michael (1999). Quasipermanent photochemical gratings in a dual use photorefractive polymer composite. *Appl. Phys. Lett. 74*, 2749.

HPCC/ESS System Performance Evaluation Project



Terrence Pratt, Senior Scientist
(pratt@cesdis.gsfc.nasa.gov)

Profile

Dr. Pratt earned B.A., M.A., and Ph.D. degrees in mathematics and computer science at the University of Texas at Austin. He is a member of the ACM, the IEEE, and SIAM. In 1972-73 he served as an ACM National Lecturer, and in 1977-78 a SIAM Visiting Lecturer. His research interests include parallel computation, programming languages, and the theory of programming.

Prior to joining CESDIS, Dr. Pratt held teaching and research positions at Michigan State University in East Lansing, the University of Texas at Austin, and the University of Virginia. At the latter he was one of the founders of the Institute for Parallel Computation and served as its first director.

During the 1980s, Dr. Pratt worked with scientists at USRA's ICASE and NASA Langley on the development of languages and environments for parallel computers. He is the author of two books: Programming Languages: Design and Implementation (Prentice-Hall, second edition, 1984) and Pascal: A New Introduction to Computer Science (Prentice-Hall, 1990).

Dr. Pratt joined CESDIS as the Associate Director in October 1992 and was appointed Acting Director in October 1993 upon the retirement of Raymond Miller. He served in that capacity until November 1994 when he left CESDIS to pursue other interests, but maintained ties with CESDIS as a consultant on high performance Fortran. He rejoined CESDIS as a Senior Scientist early in 1996.

Report

This research project is part of the NASA HPCC Earth and space science (ESS) project centered at Goddard. The ESS project funds nine "grand challenge" science teams at various universities and federal research laboratories. In addition, through a cooperative agreement with SGI/Cray, a 512 processor SGI/Cray T3E parallel computing system has been placed at Goddard to serve as a testbed system in support of the science team projects. During 1998, this system was upgraded by NASA to 1088 processors.

Each science team is responsible for developing large scale science simulation codes to run on the T3E and meet specified performance milestones (10 Gflop/sec in 1996, 50 in 1997, 100 in 1998). The codes are provided to an in-house science team at Goddard for performance verification, and ultimately the codes are submitted to the National HPCC Software Exchange for general distribution. For an overall view of the NASA HPCC/ESS project and its current status, visit the web page at http://sdcd.gsfc.nasa.gov/ESS/. For the current status of the project reported here, go to that web page and click on the "System Performance Evaluation" icon to get to the homepage for this project.

1. Research Goals

The CESDIS System Performance Evaluation Project is concerned with the large scale science simulation codes produced by the nine Grand Challenge science teams, their behavior on the massively parallel testbed computer system, and to a lesser extent their behavior on other parallel systems such as the CESDIS and NASA Beowulf systems.

Our interest is in understanding how these large science codes stress the parallel system and how the parallel system responds to these stresses. In particular, we wish to find ways to:

- Quantify the stresses produced by the science codes on the testbed hardware and software.
- Quantify the performance responses produced by the system.
- Determine the causes of the observed responses in the codes and systems.
- Use the results to improve codes and systems.
- Develop new performance evaluation and prediction methods and tools as needed.

Ultimately the goal is publication of the results of this work in various journals and conference proceedings.

2. Approach

Our approach is to work directly with the science codes as they are submitted by the science teams to meet performance milestones. We use various measurement tools to understand the static structure of each code and its dynamic behavior when executed with a typical data set (also provided by the science team). Typically, a code is "instrumented" to collect the desired statistics and timings, and then run on the testbed system using various numbers of processing nodes. The results are analyzed, and if more data are required, the instrumentation is modified and the code rerun.

The insights gained from this research on a particular code often lead to understandings about how to improve the performance of the code. These insights are fed back to the science team to aid them in further development of the code. Results may also be useful to SGI/Cray in improving their hardware and software systems, so results are often forwarded to the in-house SGI/Cray team and the in-house science team.

3. Measurements of Interest

Part of the research effort is to determine what aspects of science code structure and behavior have the greatest effect on performance. To this end, we are measuring some of the following elements in each code:

- Flops counts and rates.
- Timings and execution counts of interesting code segments.
- Data flows between code segments.
- MPI/shmem/PVM message passing and synchronization profiles.
- I/O activity profiles.
- Cache use issues.
- Storage allocation sizes and use profiles.
- Scaling with problem size and number of processors.
- Load balance.

4. Tools Used

These studies use a variety of tools for instrumenting and measuring various characteristics of the science codes and their behavior. The primary tool to date has been a software system called Godiva (GODdard Instrumentation Visualizer and Analyzer) developed by this project.

5. Current Results

All of the major results of this project are available through the project web pages (URL above). Briefly summarized, the two major results from the work during this project year are:

(1)The development of a new set of methods that allow real application codes to be used more effectively as performance benchmarks in the evaluation of large-scale parallel computer systems. The methods, and surrounding rationale for their use, are described in the report "How to Quantify an Application Code to Create a Benchmark", which is accessible through the project web pages.

To illustrate the application of the methods to a well-known example, we have used the LU benchmark from the NAS Parallel Benchmarks suite. See the report "Quantifying the NAS LU Parallel Benchmark", accessible through the project web pages, for a full description, including some surprises.

(2) Major improvements have been made in the GODIVA software system for performance measurement of large Fortran and C science codes running on big parallel machines. Composite reports showing load, load balance, and performance variations across hundreds of nodes are now easily produced. Lots of other improvements to the system have been made. The entire Godiva 4.0 Users Manual is available for download from the project web pages.

Because the complete results are available through the project web pages and may be studied more easily there, this report does not attempt to duplicate that information. Rather, the reader is advised to check directly with the web pages.

6. Status and Conclusion

This part of the evaluation project has now concluded because of the retirement of the Principal

Investigator. Although not all of the planned work was completed, two important useful products have resulted from the effort: (1) the methods for quantifying science codes to make them into useful benchmarks, and (2) the Godiva software system for performance measurement. The project may be considered a modest success.

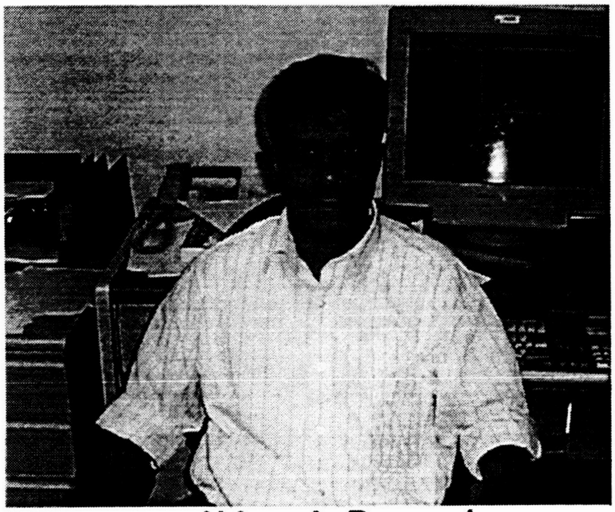
Publications

Pratt, T. (1998). Design of the GODIVA performance measurement system, in D. O'Hallaron (ed), *Fourth Workshop on Languages, Compilers, and Run-time Systems for Scalable Computers*, Pittsburgh, *Lecture Notes in Computer Science, Vol. 1511*, Springer, 219-228.

Pratt, T. (1998). Using GODIVA for data flow analysis. *Proc. SIGMETRICS Symposium on Parallel and Distributed Tools*, Welches, Oregon, ACM Press, 92-100.

Pratt, T. (1999). *Godiva 4.0 Users Manual.* CESDIS, 34pp.

Highly-parallel Integrated Virtual Environment (HIVE)



Udaya A. Ranawake
University of Maryland Baltimore County
Department of Computer Science and Electrical Engineering
(udaya@neumann.gsfc.nasa.gov)

Profile

Dr. Ranawake received a B.S degree in Electrical Engineering from University of Moratuwa, Sri Lanka in 1982, and an M.S degree in Electrical Engineering and a Ph.D degree in Computer Engineering from Oregon State University in 1987 and 1992 respectively. Prior to joining CESDIS on a subcontract with the Department of Computer Science and Electrical Engineering at University of Maryland Baltimore County, he was a senior member of the technical staff at Hughes STX Corporation where he was the task leader for massively parallel research at NASA GSFC. His research interests are algorithms for scientific computation, parallel and distributed computing, computer architecture, and computer networks. Dr. Ranawake is a member of the IEEE.

Report

1. Introduction

The rapid increase in performance of commodity microprocessors and networking hardware has provided the opportunity for exploring the potential of Pile-of-PCs (PoPC) as a low cost alternative to high end supercomputers in scientific computations. The PoPC model is used to describe a loose ensemble or cluster of PCs applied in concert to a single problem. It is similar to network of workstations (NOW) but emphasizes the use of mass market commodity components, dedicated processors, and a private system area network (SAN). The hardware components used by these systems benefit from declining prices resulting from heavy competition and mass production. This approach also permits technology tracking, allowing computing systems to be acquired with the best, most recent technology and at the lowest price. As the systems are not preconfigured by a vendor, individual systems could be configured to suit user needs. Also, the free software base available for these systems is quite robust and as efficient as commercial grade software. In early 1994, a project based on the PoPC model was initiated at NASA GSFC and is called the Beowulf project.

2. Overview of the Hive

The HIVE is a computer based on the PoPC model consisting of 64 nodes. The HIVE project's goal is to produce an inexpensive high performance parallel computer that is reliable and easy to use. The primary applications on the HIVE are earth science data manipulation, space data image restoration, ocean and atmosphere modeling, and other related applications.

The HIVE consists of dual 200 Mhz Pentium Pro rack mounted PCs for a total of 128 processors. Two additional PCs are used as hosts: a system host and a user host. The purpose of the system host is to maintain and monitor the HIVE. The user host is intended for application development and job submission to the HIVE. The nodes are interconnected with a 100 MHz full duplex fast Ethernet switch. The HIVE consists of 28 Gbytes of RAM and 900 Gbytes of disk storage distributed across the nodes.

3. Accomplishments

3.1 Upgrading and Software Configuration of the HIVE

As the co-investigator of the HIVE project, I played an active role in upgrading and software configuration of the HIVE. The memory on each node was increased from 64 MBytes to 448 MBytes and the disk capacity on each node was increased from 2.5 GBytes to 14 GBytes. Also, the 5 100 MHz full-duplex fast ethernet switches were replaced by a single fast ethernet switch resulting in significant improvements in the communication bandwidth. The system has been highly reliable, and has experienced only a few node crashes. The HIVE software environment includes programming languages such as C, C++ and aCe and interprocess communication software packages such as PVM, MPI and BSP.

3.2 The Bview Software Tool

I implemented a new version of the Bview software tool. This new version fixed some problems of the previous version and also added some new functionality to the program. Also, a paper on Bview was presented at a conference.

Bview is a software tool that displays the CPU and memory usage statistics of all the nodes of a cluster of PCs. The information is displayed in the form of a bar chart with one entry for each node in the system. The delay between screen updates is set using the bmod software tool by the super user. One may determine which bar belongs to which node by placing the cursor over that bar. This will cause a window to appear which will contain the name of the node. The status window also allows one to open a shell window on any node by clicking on its respective bar. Commands such as top may be executed within this window to obtain a more detailed view of the resource usage on a node.

Bview has two modes of operations - the normal mode and the burst mode. Bview normally operates in the normal mode; a user can select the burst mode using the menu to obtain a faster screen update. The delay between the screen updates (in normal and the burst modes) and the duration of the burst mode can be changed using the bmod software tool.

A menu provides users with the following options: change color, save current settings as the default, switch to burst mode, view the current values for the delay (in normal and burst modes) and the duration of burst mode, and quit. The current settings could also be saved as the defaults using <ctrl S> on the keyboard when the mouse cursor is on the bview window.

The heart of the 'bview' software tool is a daemon called 'bstat' that runs on each node of the PC cluster to collect statistics on CPU and memory usage. The communication between the daemon processes is done via sockets using algorithms that employ a logarithmic number of communication steps. Studies on the 64 node HIVE computer have shown that the 'bstat' daemon incurs negligible overhead when collecting statistics at 1 second intervals. The user interface part of the 'bview' software tool is implemented using TCL/TK. This software is available as part of the HIVE software archive under http://newton.gsfc.nasa.gov/thehive.

3.3 Application Development and Benchmarking

I evaluated the performance of MPICH, LAM, and PVM communication software packages on the Hive. The evaluation was performed using the MPBench benchmark suite with minor modifications. The performance of the functions such as round trip, gap time, bandwidth, broadcast, reduce, allreduce and barrier were evaluated. I also supervised a summer student who evaluated the performance of NAS parallel benchmarks on the HIVE. The results of these performance studies are available at http://newton,gsfc.nasa.gov/thehive/thehive_dir/performance.html

I assisted other users in porting, implementing and optimizing earth and space science applications for the HIVE. One application that was ported to the HIVE that delivers good performance is the MM5 code - a limited area weather model designed to simulate mesoscale and regional-scale atmospheric circulation. A second application that delivers good performance is the implementation of a hierarchical image segmentation algorithm that segments images by region growing and spectral clustering with natural convergence criteria.

3.4 Performance Evaluation of Alternate Networking Hardware

Myrinet is a cost-effective, high-performance, packet-communication and switching technology that is widely used to interconnect clusters of workstations. I built a 4 node myrinet based PC cluster in order to study their suitability for communication intensive applications such as fast fourier transforms (fft). The cluster consists of 166 MHz dual pentium pro processors with 128M main memory. The time for a 512x512 complex fft on 8 processors is 58 milli-seconds (with one matrix transpose; i.e when data is out-of-place) and 90 milli-seconds (with two matrix transposes; i.e when data is in-place). In contrast, a 512x512 complex fft on a MasPar MP-2 takes 25 milliseconds (with data in-place). Therefore, the myrinet based pc cluster gives reasonable performance on ffts and have a superior price performance ratio compared to the MasPar MP2. The mpi based fft implementation was adapted from the publicly available fftw package from MIT.

Publications

Ranawake, Udaya A. and Dorband, John (1999). BVIEW: A Tool for Monitoring Distributed Systems. *Lecture Notes in Computer Science 1593 (Proceedings of HPCN Europe' 1999),* pp. 1167-1170.

Dorband, John, Kouatchou, Jules, Michalakes, John and Ranawake, Udaya (1999). Implementing MM5 on NASA Goddard Space Flight Center Computing Systems: a Performance Study. *Proceedings Frontiers'99,* pp. 200-207.

Ranawake, U., Dorband, J., Fryxell, B., Ridge, D., Hendriks, E., Becker, D. and Merkey, P. Achieving Ten Gflops on PC Clusters: A Case Study. *USRA/CESDIS Technical Report TR-98-219*, 1998.

Technical support for Emergency Management and Regional Applications Center Development

Fran Stetina
Fran Stetina and Associates
(stetina@gsti.com)

Objective

This task provides technical consulting support to develop an emergency management applications and technology transfer applications program. Program elements will include: hyper spectral aircraft and MODIS spacecraft remote sensing instruments, an Unmanned Aerial Vehicle for remote sensing, the Earth Alert and Weather Anywhere wireless information dissemination systems, new search and rescue concepts and the use of data and products from the Regional Applications Centers to provide information for man-made and natural hazard situation support.

Background

NASA/GSFC has recently signed a Memorandum of understanding with Federal Emergency Management Agency to transfer GSFC technology to the Emergency Management community. To implement this agreement, GSFC proposes to develop an Emergency Management Application Program.

FEMA has identified a number of GSFC technologies which will have immediate potential in reducing loss of life by providing personal warnings to people in harms way. These include the Earth Alert and the Weather Anywhere projects

In addition, the emergency management community has identified a need for rapid and seamless information distribution to support various emergency situations, such as major forest and brush fires, environment hazards emergencies and natural disasters. Because of the local nature of many of these disasters, they have identified the Regional Applications Centers as a source of information and data products. GSFC has helped develop a number of Regional Applications Center which are distributed throughout the nation.

These Centers would contribute to the efficient and effective utilization of human and natural resources and the development of an information infrastructure to support knowledgeable decision making. Such an infrastructure must not only gather and store data, but it must contain sufficient processing power and intelligence to produce useful output products. The system must facilitate rapid retrieval and distribution of information so that decision making can be made based on objective criteria using expert knowledge and simple visualization techniques. This philosophy requires a systems design approach which emphasizes integration, automation, user friendly interfaces and thorough understanding of the users requirements. Since many regional issues require high resolution hyper-spectral data, emphasis will be placed on exploiting both aircraft hyper-spectral instruments and the space-borne MODIS multispectral instrument.

Implementation of systems with the above mentioned features is based on 10 years of project management experience for NASA/GSFC in implementation of satellite weather receiving systems, ground processing and; specifically it includes the development of a modular system concept called SAMS, Spatial Analysis & Modeling System. The SAMS system has been defined as a potential model for the development of the Regional Applications Centers.

One of the key components of such a system is a real time, direct readout capability. Thus, the design and development of Regional Environmental and Technology Center concept (Regional Application Center), has been defined as an important objective of NASA's Earth Science Enterprise.

Scope

The activities to be undertaken under this task include hardware and software system design which are required to enhance the existing prototype Regional Application Centers. In addition, requirements analysis, user interfaces, and user specific products definitions and descriptions are required.

The task also includes the development of an emergency management program plan and an implementation plan for the various projects which are defined within this program.

Included in this concept is the need to develop a core EOS direct readout capability and general support of end-to-end system software to provide EOS core instrument algorithms and basic mission products. The system concept should include an archiving and distribution capability. In addition, strategies should be developed to test EOS direct readout system components and concepts in an operational environment. This includes the use of aircraft high spatial and spectral resolution instruments to support algorithm development and evaluation, integration of in situ measurements to validate remote sensing measurements, and the integration of a Geographic Information System. The system should include the design of a local user analysis system to interface with the Regional Application Center.

Task Elements

- Provide expert advice to determine user requirements for EOS Direct Readout core instrument algorithms and products.

- Assist in working with EOS science working groups and instrument scientists to define status and availability of algorithms.

- Determine product requirements for international science and operational users.
- Develop project plans to utilize hyper-spectral instruments to facilitate the development of regional EOS MODIS algorithms.
- Assist in defining Earth Science core algorithm processing capability for a direct readout facility.
- Provide expert advise in defining EOS direct readout system concepts, and define end-to-end system components and functions.
- Develop operational scenarios for direct readout system and its interfaces with the GSFC EOSDIS.
- Provide expert advise in the development of strategies to develop and test various components of the Regional Applications Center system using existing operational facilities.
- Determine weather product requirements for various applications which will be implemented at Regional Applications Centers for both operational and research users.
- Assist in the development of a field project to demonstrate the Weather Anywhere system.
- Assist in the development of a NASA Applied Emergency Management Program and define an implementation plan for a pilot Regional Disaster Management and Communications Information Center.
- Perform the functions of liaison between the Pacific Disaster Center and the NASA Hawaii Regional Application Center.
- Provide continuing support in obtaining funding for the Earth Alert and Weather Anywhere systems development and develop joint agency demonstration field tests of these systems.
- Support the system definition of a Search and Rescue Communications Information System.
- Represent the NASA/GSFC Regional Data Center manager in meetings and conferences as required.
- Define potential international site locations and projects which may become the basis for Regional Calibration and Validation Centers.
- Assist in the development of an implementation plan for the use of Unmanned Aerial Vehicles to support Earth Science Enterprise regional algorithm development and product validation
- Assist in the implementation of the NASA/GSFC FEMA Memorandum of Understanding.

The Earth Alert personal warning system has been defined as a potential important technology which has significant value to the Emergency Management Community and especially to the Hawaii Pacific Disaster Center. A number of activities relating to bringing this technology to a successful commercial product line.and introduction of this capability to the Hawaii Civil Defense and to FEMA have been undertaken under this task.

Accomplishments

Under the activities of this task, a Memorandum of Understanding between GSFC & FEMA has been implemented. A number of papers and presentations have been made to introduce the Earth Alert System to the Emergency Management Community. In addition I have participated in a number of FEMA sponsored partnership workshops:

- FEMA workshop at ARGONE LABS, Chicago, IL July 98
- Detailed meetings and discussions were held with the Maryland EOC to develop a field test program for the Earth Alert Project. Similar discussions were held with the City of Houston EOC to define an Earth Alert Project experiment to demonstrate how the system would be used in a hazardous chemical spill.
- Project impact Workshops March 99, May 99
- FEMA workshop at OAK RIDGE Lab, Gatlinberg,TN May 99
- Fire Ass. technical meeting Emmitsberg, MD June 99

I have helped develop a straw man strategic plan to fully implement the NASA-FEMA MOU. This plan is being used to define a NASA initiative for enhancing an applications outreach project to support the NASA Headquarters Natural Hazards Program.

A plan has been developed to extend this warning system to a more general information dissemination system called " Weather Anywhere." Details of the system were presented at the Annual Air Traffic Control Ass. Conference in Atlantic City in Nov. 98.

A plan is being developed to implement various GSFC information system technologies to support the Emergency Management Community. A preliminary outline of the plan is attached as Appendix A of this report.

A study has been completed to define how the Regional Application Centers can be utilized to support Emergency Management Requirements. A copy of the study can be obtained from the Earth Alert Project Manager: Fred Schamann NASA/GSFC code 933.

Various low cost aircraft hyper-spectral instruments have been identified and are under field investigations to provide information which will be useful in developing EOS direct readout data products for regional applications. A Code 935 VIFIS wedge spectrometer instrument has been utilized by the Resource 21 project to investigate the use of hyper-spectral information for commercial applications.

- I participated in an IGARSS workshop in Seattle, Wash, June 98
- I presented a paper at the SPIE conference in Barcelona, Spain, Sept. 98
- I presented two papers at the European Workshop on Hyper spectral Imaging in Zurich, Switzerland, Oct. 98

Two Pilot EOS direct readout satellite stations are being implemented at GSFC for future use at RAC's. I have participated in the acceptance test and evaluation of these ground receiving stations.

- Factory visit to Charleston, N.C. was conducted in Feb. 99 to review system design of a low cost EOS direct read out system.

ADMINISRATION TEAM

Chang-Hong Chien, Systems Administrator
L'Tanya Clark, Administrative Assistant 3 (Financial)
Lakeena Courtney, Administrative Assistant 1 (at time of submission)
Georgia Flanagan, Administrative Assistant 3 (Conference Management)
Michele Meyett, Administrative Assistant 2 (Web Site Administration, Database Management, Presentation Graphics, Desktop Publishing)
Dawn Segura, Promoted to Administrative Assistant 3 (Procurement Specialist)
Yolanda Smith, Administrative Assistant 2 (Event/Visitor Support, Human Resources)

This branch is responsible for supporting the CESDIS Director, Senior and Staff Scientists, Technical Specialists, funded project personnel and graduate students, and USRA's corporate office. Branch personnel:

- Serve as the liaison among funded research personnel, NASA scientific and administrative personnel, and USRA accounting and procurement personnel,
- Monitor subcontracts and consulting agreements,
- Monitor the contract's Small and Small/Disadvantaged Business Plan,
- Prepare and monitor task budgets,
- Prepare contract reports,
- Obtain Contracting Officer permission for foreign travel by staff and university scientists,
- Obtain Contracting Officer permission for equipment purchases with contract funds and report purchases to Goddard's property personnel,

- Assist with conference planning and provide on-site support at conference, workshop, and seminar locations,
- Assist foreign national visitors in gaining access to Goddard,
- Provide peer review support to NASA program personnel for proposals submitted in response to NASA Research Announcements and Cooperative Agreement Notices,
- Maintain CESDIS Web site,
- Provide desktop publishing assistance for paper preparation, the CESDIS newsletter, and presentation graphics,
- Make travel arrangements and provide assistance with travel voucher completion,
- Perform functions of remote site data entry for USRA's centralized accounting system including payroll, purchasing, and accounts payable.

ADMINISTRATION ACTIVITIES

Seminar Series

CESDIS sponsors seminars by visiting scientists from universities, government laboratories, and the public sector. These presentations are open to everyone at Goddard as well as interested off-site attendees. Announcements of speakers and dates are posted on the CESDIS Website. Seminar presentations during this reporting year are listed below. Abstracts appear in the Outreach and Education Section.

- Dr. S Rao Kosaraju. Johns Hopkins University. *On the Optimal Split Tree Problem.*
- Dr. Samuel J. Lomonaco Jr. University of Maryland Baltimore County. *An Overview of Quantum Computation: Concept and Intuition.*
- Dr. Peter Norris. National Institute of Water and Atmospheric Research. *A study of nocturnal marine stratocumulus development using Lagrangian (particle-based) large-eddy simulation .*
- Dr. Patrick Kinney. Columbia University. *Oaone and Epidemiological Studies.*
- Irene Qualters. *Pratical Approach to Guiding Large, Living Software Projects.*
- Prof. Eero P. Simoncelli. New York University. *Rotation and Translation-invariant Image Representation.*
- Prof. Anupam Joshi. University of Maryland Baltimore County. *Data Mining, Web Mining, and Computational Intelligence.*
- Dr. Vipin Kumar. University of Minnesota. *Data Mining in Very Large Dimensional Data Sets.*
- I. Michael Navon. Florida State University. *Strategies for four-dimensional variational data assimilation using the FSU Global Spectral Model with its full physics adjoint.*
- Dr. Fedor Mesinger. National Center for Environmental Function. *The Eta Model: Design, Performance, Some Conclusions, Future.*
- Dr. Ron Minnich. David Sranoff Labs. *Beowulf and other Clusters.*
- Dr. Yvan G. Leclerc. SRI International Artificial Intelligence Center. *Visualizing the Earth using TerraVision II.*
- William E. Johnston. NASA Ames and Lawrence Berkeley National Laboratory. *The Information Power Grid Project: Research, Development, and Testbeds for High Performance, Widely Distrbuted, Collaborative Computing and Information Systems Supporting Science and Engineering.*
- Mark Lucas. ImageLinks, Inc. *Automated Image Registration with Parameter Adjustment.*
- Robert B. Ross. Clemson University. *Message Passing and Parallel File Systems for Beowulf Machines.*
- Mr. John Levesque. Advanced Computing Technology Center. *The Advanced Computing Technology Center IBM Watson Research Laboratory.*

- Robert E. McGrath. University of Illinois, Urbana-Champaign. *Integrating Scientific Datasets and Digital Libraries.*
- Dr. Takashi Iida. Communications Research Laboratory of Ministry of the Posts & Telecommunications. *Japan Gigabit Satellite Project.*
- Rick Paxman. ERIM International. *Phase-Diversity Technology: Wavefront Sensing and Imaging.*
- Jacob B. Khurgin. *Generation of Terahertz radiation: comparative analysis.*
- David Brunnell. *Cinebase.*
- Daniel C. Edelson. Northwestern University. *Adapting Scientists' Investigation Tools for Inquiry Learners: A Case-Study of Visualization in Earth Systems Science.*
- Charles L. Seitz. Myricom Inc. *Myrinet -- Scalable Cluster Interconnect.*

CESDIS Science Council

The CESDIS Science Council met on December 7, 1998 at Goddard. Presentations on work-in-progress were given by Yelena Yesha, Harold Stone who spoke of his collaborative work with Jacqueline Le Moigne who could not be present, Don Becker, Richard Lyon, Nathan Netanyahu, Susan Hoban, and Kostas Kalpakis. A portion of the afternoon was devoted to an open discussion of the future of CESDIS by interested participants since the second 5-year contract was due to expire on July 5, 1998. (Ultimately the existing contract was extended for two years through July 5, 2000.)

The next regularly scheduled meeting of the CESDIS Science Council will be in the Fall of 1999.

NASA Summer School for High Performance Computational Physics

The NASA Goddard Space Flight Center's Earth and Space Data Computing Division (ESDCD) and the Universities Space Research Association solicited applications from graduate students to participate in an intensive lecture series in computational phycis during a three-week period. The ESDCD provided comprehensive research and development support in data handling and computing for NASA Earth and Space Science Research Programs. Resident facilities included a 512-processor Cray T3E, a Cray J90 cluster composed of three 32-processor Cray J90 systems, and a MasPar MP-2/MP-1 cluster. The program stemed from ongoing activities that reflected NASA's desire to help train the next generation of physicist in the development of computational techniques and algorithms for scalable parallel copmuters in support of the Federal High Performance Computing Communications Program.

Visiting Student Enrichment Program (VSEP)

The VSEP program was jointly sponsored by Goddard Space Flight Center's Earth and Space Data Computing Division and other participating GSFC organizations.

The Visiting Student Enrichment Program offered students summer employment with Universities Space Research Association (USRA), working with NASA/Goddard Space Flight Center's (GSFC) scientists. Student projects included simulating a neural network, preparing image analysis algorithms on supercomputers, developing computational science applications, and creating interactive World Wide Web sites.

Director's Special Symposium

The Director's Special Symposium was held at Goddard on January 13, 1999. The topic of discussion was Applications of Remote Sensing: Fire Detection and Modeling.

The format of the Symposium will be informal round-table discussion of research and development that may contribute to the detection and modeling of forest, grassland and bush fires using remote sensing data.

Workshop on The Roles of Computer Simulation

This workshop was held in recognition of CESDIS 10th Anniversary at Goddard on January 20-21, 1999.

The 7th Symposium on the Frontiers of Massively Parallel Computation

This conference was sponsored by the IEEE Computer Society in cooperation with NASA Goddard Space Flight Center and USRA/CESDIS, held in Annapolis, Maryland on February 21-25, 1999. The conference provided a major forum for exploring technical issues that are driving the outer boundaries of effective high performance computing. The series of symposia is one of the prinicpal meetings for presenting new and original research results extending the threshold of computational capability through advances in hardware, software, methods, and technology. The spectrum of fields addressed by the conference included applications and algorithms, system software and languages, component technologies, and system architectures.

Advances in Digital Libraries

This conference was held in Baltimore, Maryland on May 19-20, 1999. The conference shared and disseminated information about important current issues concerning digital library research and technology. The goal was achieved by means of research papers, invited talks, workshops, and panels involving leading experts, as well as through demonstrations of innovative and prototype technologies. The conference had the additional goal of indicating the importance of applications of digital library technologies in the public and private sectors of the economy.

NASA SUMMER SCHOOL FOR HIGH PERFORMANCE COMPUTATIONAL PHYSICS
Summer 1998

The NASA Goddard Space Flight Center's Earth and Space Data Computing Division (ESDCD) and the Universities Space Research Association solicited applications from qualified graduate students to participate in an intensive lecture series in computational physics during the three-week period July 13 - 31, 1998. The ESDCD provided comprehensive research and development support in data handling and computing for NASA Earth and space science research programs. Resident facilities included a 512-processor Cray T3E, a Cray J90 cluster composed of three 32-processor Cray J90 systems, and a MasPar MP-2/MP-1 cluster. This program stemmed from ongoing activities that reflected NASA's desire to help train the next generation of physicists in the development of computational techniques and algorithms for scalable parallel computers in support of the Federal High Performance Computing Communications Program.

Approximately 15 students were selected to participate in the three-week program. Students were given hands-on computer training and small group interaction experience. Experienced computational scientists presented series of lectures on advanced topics in computational physics, with emphasis on computational fluid dynamics and particle methods. Cray Research presented lectures on developing software for their massively parallel architectures. Both the Cray T3E and the MasPar MP-2/MP-1 cluster were available for use by the students. At the end of the program, students were required to present a 15-minute summary of what they learned and how it relates to their respective fields of study.

The program aimed to attract Ph.D. students in the Earth and space science disciplines whose present or future research requires large-scale numerical modeling on massively parallel architectures. Eligibility was normally limited to those Earth and space science students who were enrolled in U.S. universities and who passed their Ph.D. qualifying exams. Because of NASA GSFC security regulations, citizens of certain prescribed nations were ineligible.

Application materials included:

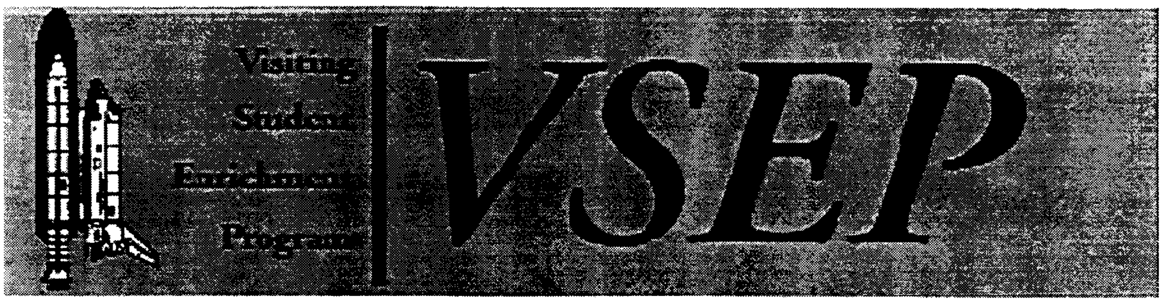
1. a cover letter explaining his/her interest in the program and how his/her research may benefit from their participation;
2. area of research and thesis title;
3. a statement of career objectives and goals;
4. a description of relevant work experience;
5. cirriculum vitae or resume with publication list;
6. current G.P.A.;
7. two letters of reference;
8. academic transcripts showing two full years of work; and
9. a statement of citizenship and visa status.

Students received a per diem and were reimbursed for domestic transportation to and from Greenbelt, Maryland. Students were housed near the Goddard Space Flight Center, and transportation to and from Goddard each day was provided. Applications were received before February 13, 1998. There were no formal application materials. Selection announcements were planned by March 6, 1998. All application information was directed to: Georgia L. Flanagan, Program Coordinator, USRA/HPCP, Code 930.5, NASA Goddard Space Flight Center, Greenbelt, MD 20771. (301) 286-2080, georgia@cesdis.usra.edu.

Launch Your Future at NASA

VISITING STUDENT ENRICHMENT PROGRAM

1999



Jointly sponsored by Goddard Space Flight Center's Earth & Space Data Computing Division and other participating GSFC organizations.

The Visiting Student Enrichment Program (VSEP) offered students summer employment with the Universities Space Research Association (USRA), working with NASA/Goddard Space Flight Center's (GSFC)
scientists. Student projects included simulating a neural network, preparing image analysis algorithms on supercomputers, developing computational science applications, and creating interactive World Wide Web sites.

Project experiences were available from **June 7 to August 13, 1999**, (high school students may start/stop 1-2 weeks later) at GSFC in Greenbelt, MD: The first, the individual research experience, matched one student with a staff member as a mentor to work on a project. The second, the group research experience, placed up to 6 students in a team that worked on a project under the supervision of a staff member. Both paths provided opportunities to work with scientists and professionals at a world-class facility while offering a meaningful work experience primarily focused on computer science or the application of computers to solve problems in other sciences. VSEP also offered field trips and lectures to broaden appreciation for the GSFC mission and activities.

Where Might I Be Working?

Organizations that participated in the past included:

- The Scientific Computing Facility provided scientists access to advanced computers like a

Cray T3E, three Cray J90's, a Convex C3830, and the world's largest Convex/UniTree mass storage system, as well as a visualization studio. This enabled researchers to model Earth's processes (weather, climate, and crustal dynamics), as well as space plasma (magnetosphere and solar phenomena) and astrophysical systems.

- The National Space Science Data Center served as a central repository for the enormous data bases generated by instruments aboard NASA spacecraft. Staff members developed space physics and astrophysics data systems, intelligent data systems, data visualization techniques, distributed databases, and new technologies for mass storage.

- Flight Dynamics Analysis Facility used computers to perform mission design and determine spacecraft attitude and orbit parameters. Staff members researched advanced techniques for mission support and systems engineering including state-of-the-art graphics techniques and advanced software engineering.

- The Data Systems Technology Division provided a full spectrum of hardware/software environments to support applied research and development of advanced technological solutions to operational problems. Application domains ranged from mission operations for near-Earth unmanned scientific satellites to administrative support systems.

- The Global Change Data Center provided Earth science data operations in studies of climate, oceanography, and land resources.

- Laboratory for Atmospheres researched areas such as atmospheric modeling and climate analysis in support of various Earth observing systems.

- Laboratory for Hydrospheric Processes researched a broad range of areas in the oceanic, cryospheric, and hydrologic sciences.

How Do I Qualify?

The Program was opened to full-time students in computer science, the physical sciences, and mathematics. All students were evaluated relative to their school-level peers. Participants were either U.S. citizens or foreign nationals in U.S. schools who possess a work visa.

College: Undergraduate and graduate students must have taken courses in physical and computer sciences directly related to their areas of study.

High School: Students were evaluated with emphasis on their potential and related extracurricular experiences, as well as on course work. The number of positions available were limited.

Did the Program Provide Remuneration?

Students were made full-time temporary employees of USRA, a nonprofit academic research consortium. The compensation rate was lower for high school students than for undergraduate/graduate students and was set before students were chosen. For those students not within normal commuting distance to GSFC, the program provided limited round-trip travel expenses and local housing at the University of Maryland.

How were Students Selected?

Participants were selected after a competitive review. Selection criteria included academic record, letters of reference, experience, and career goals/interest in VSEP. Funding was available for approximately 20 positions.

How Do I Apply?

There are no formal application forms. To be considered for VSEP, students sent the following application materials to USRA:

1. Full name and both current and permanent addresses with telephone numbers and email address, if available.
2. Social Security number and citizenship.
3. Grade level, GPA, and intended major.
4. Well-written statement of career goals and reasons for interest in VSEP.
5. Description of relevant experience.
6. Letters of reference (minimum of two).
7. Formal academic transcripts for at least the past 2 full academic years.
8. The path(s) for which him/her would like to be considered: Individual Research, Group Research, or Both Individual and Group Research.

Application Material was Directed To:

Visiting Student Enrichment Program
USRA
Mail Code 930
NASA/Goddard Space Flight Center
Greenbelt, Maryland 20771

Web: http://sdcd.gsfc.nasa.gov/VSEP/
Email: VSEP@cesdis.usra.edu
Telephone: 301-286-4403

Application Deadline:

Materials were received by January 25, 1999. Selection announcements were made by May 15, 1999.

Note: Transcripts and reference letters were sent directly from the academic institution to the address provided above.

Appendix A

Director's Special Symposium
"Applications of Remote Sensing: Fire Detection and Modeling"

January 13, 1999

Director's Special Symposium "Applications of Remote Sensing: Fire Detection and Modeling"

Wednesday, January 13, 1999
10 AM - 2 PM
Building 28, Room E210
NASA Goddard Space Flight Center

The format of the Symposium will be informal round-table discussion of research and development that may contribute to the detection and modeling of forest, grassland and bush fires using remote sensing data.

For information or to RSVP, please contact Georgia Flanagan, CESDIS Event Coordinator, at georgia@cesdis.gsfc.nasa.gov.

Appendix B

Workshop on "The Roles of Computer Simulation"

in recognition of

CESDIS 10th Anniversary

A Workshop on
"THE ROLES OF COMPUTER SIMULATION"
in recognition of the

CESDIS 10th Anniversary

January 20-21, 1999
NASA Goddard Space Flight Center
Building 28 Room E210

Agenda

WEDNESDAY JANUARY 20th

8:30 Registration/Coffee

9:30 - 10:00 Welcoming Remarks

Yelena Yesha , Director of CESDIS
Paul Coleman , President of USRA
Al Diaz , Director, NASA Goddard Space Flight Center

10:00 - 10:30 Keynote Address

The Future of Modeling in Space Science Simulation
Paul Fishwick , University of Florida

10:30- 10:45 BREAK

10:45 - 12:15 SESSION 1.1: SIMULATION IN EARTH SCIENCE
Chair: Yelena Yesha, CESDIS/UMBC

A Retrospective view of Simulation Studies
Milton Halem , NASA/GSFC

Virtual Reality in the Thermal Infrared for Forest and Grassland
James Smith , NASA/GSFC

Digital Earth A Virtual Representation of our Planet
Horace Mitchell , NASA/GSFC

12:15 - 1:30 LUNCH - ON YOUR OWN

1:30 - 3:00 SESSION 1.2: PARALLEL SIMULATION
Chair: Ian Akyildyz, Georgia Tech

The High Level Architecture for Simulations
Judith Dahmann, Defense Modeling and Simulation Office (DMSO), Virginia

Exploiting Temporal Uncertainty in Parallel and Distributed Simulations
Richard Fujimoto, Georgia Tech

Improving Automated Military Commanders in Distributed Battlefield Simulation
Billy Foss, Institute for Simulation and Training

3:00 - 3:30 BREAK

3:30 - 5:00 SESSION 1.3: APPLICATIONS I
Chair: Bill Hayden, NASA GSFC

The Intelligent Synthesis Environment: Engineering Design in the 21st Century
John Malone, NASA/LaRC

Simulation and Visualization of Landscape Processes
Doug Johnston, University of Illinois, Urbana-Champaign

AF-GOESpace: Space Environment Models for Acquisition, Operations, and M&S
Robert Hilmer, Air Force Research Lab

5:00 - 6:00 BREAK

6:00 CESDIS 10th Anniversary Banquet {Building 28 Atrium} Invitation Only

Opening remarks: **Ray Miller**, former Director of CESIDS
Banquet speaker: **Lee Holcomb**, NASA Chief Information Officer

THURSDAY JANUARY 21st

8:00 Coffee

8:30 - 9:00 Keynote Address

Simulation: The Third Leg of Science
David Nicol, Dartmouth University

9:00 - 10:30 SESSION 2.1: APPLICATIONS II
Chair: Yelena Yesha, CESIDS/UMBC

Design and Simulation of an Autonomous On-Board Optical Control System for the Next Generation Space Telescope
Rick Lyon , CESDIS/UMBC

Virtual Petaflops in Cosmology and Cosmogony
George Lake , University of Washington

Physically Accurate Visualization and Simulation of Chemical Biological Agents in the Lower Atmosphere
Patti Gillespie , Army Research Lab

10:30 - 11:00 BREAK

11:00 - 12:00 SESSION 2.2: APPLICATIONS III
Chair: George Lake, CESDIS/Univ. of Washington

DPAT: A Fast Time Parallel Simulation for Aviation Applications
Fred Wieland , MITRE

Using Simulation with Genetics-Based Maching Learning
Bruce Dike , Boeing Corp.

12:00 - 1:30 LUNCH - ON YOUR OWN

1:30 - 3:00 SESSION 2.3: DISTRIBUTED SIMULATION
Chair: Susan Hoban, CESDIS/UMBC

Distributed Fault Tolerance Databases for Distributed Simulation
Chris Wallace , Lockheed Martin Information Systems

New Simulation Paradigms for Distributed Intelligent Control of Large Scale Discrete Event Systems
Wayne Davis , University of Illinois, Urbana-Champaign

Modeling and Simulation of Global INTERNET
Andy Ogielski , Rutgers University

3:00 - 3:30 BREAK

3:30 - 4:30 SESSION 2.4: APPLICATIONS IV
Chair: Ian Akyildiz, Georgia Tech

Simulation of Atmospheric Effects on Acoustic Propagation and Detection
Keith Wilson , Army Research Lab

Computer Statistics and Simulation in the Next Generation
William Conley , University of Wisconsin at Green Bay

Appendix C

The 7th Symposium on the Frontiers of Massively Parallel Computation

February 21 - 25, 1999

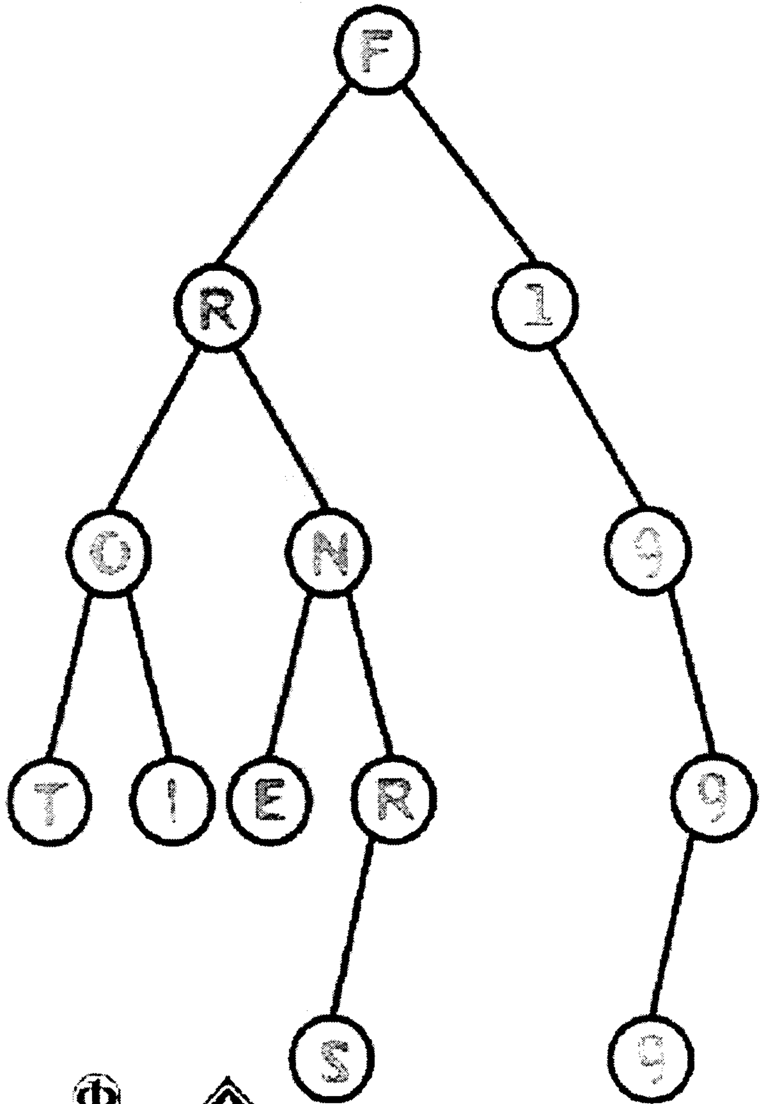
Final Program

The 7th Symposium FRONTIERS Massively

on the 1999 Parallel

Annapolis, Maryland

Frontiers of February 21 - 25 Computation



Sponsored by:
IEEE Computer Society

COMPUTER SOCIETY

In cooperation with
NASA Goddard Space Flight Center
USRA/CESDIS

Committee Members

General Chair
Charles Weems
University of Massachusetts
weems@cs.umass.edu

Program Chair
Ian Foster
Argonne National Laboratory
itf@mcs.anl.gov

General Co-Chair
Bill Carlson
Institute for Defense Analyses
wwc@super.org

Program Vice Chairs
The PetaaFlops Frontier:
Peter Kogge
University of Notre Dame

Today's Teraflop Systems:
Dave Nowak
Lawrence Livermore National Laboratory

Massive Parallelism for the Masses:
Mike Smith
Harvard University

Conference Support:
Georgia Flanagan
USRA/CESDIS

Program Committee
D. Bader
University of New Mexico

D. Bailey
National Energy Research Scientific Computing Center

K. Batcher
Kent State University

F. Berman
University of California, San Diego

J. Brockman
University of Notre Dame

D. Buell
Institute for Defense Analyses

D. Chen
University of Notre Dame

D. Elliot
University of Alberta (Canada)

G. Fox
Syracuse University

L. Freitag
Argonne National Laboratory

D. Greenberg
Institute for Defense Analyses

W. Gropp
Argonne National Laboratory

M. Heath
University of Illinois, Urbana-Champaign

J. Ja'Ja'
University of Maryland, College Park

C. Koelbel
Rice University

B. Lim
IBM Watson

R. Lucas
NERSC

Workshops/Tutorials

Sunday 21st February

Workshop A: Scientific and Engineering Computing with Applications
Part 1, 1 - 5pm; Tiaruo Wang

Workshop C: Innovations in Quantum Computation and Communications
1 - 5pm; John Thorp

Monday 22nd February

Workshop A: Scientific and Engineering Computing with Applications
Part 2, 8:30am - 12:00 noon; Tiaruo Wang

Workshop D: Third Workshop on Petaflops
8:30am - 5pm; Guang Gao

Workshop E: Reconfigurable Computing—Adaptive Computing Technology
1 - 5pm; John Schewel

Technical Program

Tuesday 23rd February

9:00-10:00 Keynote Address:
Ken Kennedy, Professor Of Computational Engineering, Rice University
"Future Investment in Information Technology Research: Report of the President's Information Technology Advisory Committee"

10:30-12:00 Technical presentations

Session 1: Architecture

10:30-11:00 Scalability Analysis of Multidimensional Wavefront Algorithms on Large-Scale SMP Clusters
Adolfy Hoisie, Olaf Lubeck, Harvey Wasserman
Los Alamos National Laboratory

11:00-11:30 A System for Evaluating Performance and Cost of SIMD Array Designs
Martin Herbordt, Jade Cravy, Renoy Sam, Owais Kidwai, Calvin Lin
University of Houston

11:30-12:00 Design Trade-Offs of Low-cost Multicomputer Networks
Martin Herbordt,
University of Houston,
Kurt Olin, Harry Le
Compaq Computer Corporation

Session 2: Software

10:30-11:00 The Cactus Framework for Computational Astrophysics
Ed Seidel
Max Planck Institute, Germany

11:00-11:30 The PETSc Library for Scientific Software
Lois Curfman McInnes
Argonne National Laboratory

11:30-12:00 The POOMA Object-Oriented Framework
John Reynders
Los Alamos National Laboratory

13:30-15:00 Technical Presentations

Session 3: Distributed Computation

13:30-14:00 Distributed Applet-based Certifiable Processing in Client/Server Environments
Gerald Masson, Hongxia Jin, Gregory Sullivan
Johns Hopkins University

14:00-14:30 Latency tolerant Algorithms for WAN Based Workstation Clusters
Mark Clement, Bernd Helzer, Quinn Snell
Brigham Young University

14:30-15:00 Large-Scale Distributed Computational Fluid Dynamics on the Information Power Grid using Globus
Stephen Barnard, Rupak Biswas, Subhash Saini, Robert Van der Wijngaart, Maurice Yarrow, Lou Zechtzer
NASA Ames Research Center
Ian Foster, Olle Larsson
Argonne National Laboratory

Session 4: ASCI and Data Visualization Corridors

13:30-14:00 Early Results from the ASCI Program
David Nowak
Lawrence Livermore National Laboratory

14:00-14:30 Data-Visualization Corridors
Rick Stevens
Argonne National Laboratory

14:30-15:00 Distance Corridors
Carl Kesselman
USC Information Sciences Institute

15:30-17:00 Technical Presentations

Session 5: Data parallelism

15:30-16:00 A Framework for Generating Task Parallel Programs
Ursula Fissgus, Thomas Rauber
University Halle-Wittenberg,
Gudula Ruenger
University Leipzig

16:00-16:30 HPF Implementation of ARC3D
Michael Frumkin, Jerry Yan
NASA Ames Research Center

16:30-17:00 Packing/Unpacking Information Generation for Efficient Generalized kr -> and r ->kr Array Distribution
Yeh-Ching Chung, Ching-Hsien Hsu
Feng-Chia University

Session 6: Systems

15:30-16:00 Efficient VLSI Layouts of Hypercubic Networks
Chi-Hsiang Yeh, Manos Varvarigos, Behrooz Parhami
University of California, Santa Barbara

16:00-16:30 Adapting to Load on Workstation Clusters
Robert Brunner, Laxmikant Kale
University of Illinois at Urbana-Champaign

16:30-17:00 Parallel Simulation of Two-Phase Flow Problems Using the Finite Element Method
Shahrouz Aliabadi, Khalil Shujaee
Clark Atlanta University
Tayfun Tezduyar
Rice University

17:00-18:00 Panel: Whatever Happened to SIMD?
John Dorband, *Computer Scientist, NASA Goddard Space Flight Center*; Leo Irakliotis, *Professor of Computer Science, University of Chicago*; Stewart Reddaway, *Chief Techical Officer, Cambridge Parallel Processing*; Charles Weems, *Professor of Computer Science, University of Massachusetts;* Stewart Reddaway, *Chief Scientist, Cambridge Parallel Processing*

18:00 Reception

Wednesday 24th February

9:00-10:00 Keynote Address:
Gil Weigand, Deputy Assistant Secretary for STRA, Department of Energy

10:30-12:00 Technical presentations

Session 7: Applications

10:30-11:00 A Data-Parallel Algorithm for Iterative Tomographic Image Reconstruction
Calvin Johnson
National Institutes of Health
Ariela Sofer
George Mason University

11:00-11:30 Parallel Rendering of 3D AMR Data on the SGI/Cray T3E
Kwan-Liu Ma
NASA Langley Research Center

11:30-12:00 A Recursive PVM Implementation of an Image Segmentation Algorithm with Performance Results Comparing the HIVE and Cray T3E
James Tilton
NASA Goddard Space Flight Center

Session 8: Performance Modeling

10:30-11:00 Performance Engineering: An Integrated Approach
Dan Reed
University of Illinois at Urbana-Champaign

11:00-11:30 POEMS - End-to-End Peformance Models for Parallel and Distributed Systems
J.C. Browne
University of Texas at Austin

11:30-12:00 Application Driven Performance Extrapolation
Joel Saltz
University of Maryland

13:30-15:30 Technical Presentations

Session 9 Systems

13:30-14:00 MPI: The Only Progamming Model for Managing Memory
William Gropp
Argonne National Laboratory

14:00-14:30 Distributed Control Parallelism for High Speed Civil Transport Multi-Disciplinary Optimization
Layne Watson, Denitza Krasteva, Chuck Baker, William Mason
Bernard Grossman
Virgina Tech
Rafael Haftka
University of Florida

14:30-15:00 Interprocedural Communication Optimizations for Message Passing Architectures
Gagan Agrawal
University of Delaware

15:00-15:30 Data Sieving and Collective I/O in ROMIO
Rajeev Thakur, William Gropp, Ewing Lusk
Argonne National Laboratory

Session 10: Application Optimization

13:30-14:00 The Parallelization of a Highway Traffic Flow Simulation
Charles Johnston
Concurrent Computer Corporation,
Anthony Chronopoulos
University of Texas at San Antonio

14:00-14:30 Implementing MM5 on NASA Goddard Space Flight Center Computing System: a Performance Study
John Dorband
NASA Goddard Space Flight Center
Udaya Ranawake
CESDIS/UMBC
Jules Kouatchou
Morgan State University
John Michalakes
Argonne National Laboratory

A. Malagoli
University of Chicago

R. Martino
National Institutes of Health

K. Murakami
Kyushu University

M. O'Keefe
University of Minnesota

S. Ranka
University of Florida

D. Reed
University of Illinois, Urbana-Champaign

J. Reynders
Los Alamos National Laboratory

J. Saltz
University of Maryland, College Park

E. Sha
University of Notre Dame

H. J. Siegel
Purdue University

M. Snir
IBM TJ Watson

T. Sterling
CalTech/JPL

J. Torrellas
University of Illinois, Urbana-Champaign

A. Veidenbaum
University of Illinois at Chicago

P. Wang
George Mason University

Steering Committee

Jim Fischer, Chair
NASA Goddard Space Flight Center

Larry Davis
University of Maryland

Judy Devaney
National Institute of Standards and Technology

Jack Dongarra
Oak Ridge National Laboratory

John Dorband
NASA Goddard Space Flight Center

Milt Halem
NASA Goddard Space Flight Center

Paul Messina
CalTech

Merrill Patrick
National Science Foundation

David Schaefer
George Mason University

Paul Schneck
MRJ Technology Solutions

H. J Siegel
Purdue University

Thomas Sterling
CalTech and JPL

Francis Sullivan
Center of Computing Sciences

Pearl Wang
George Mason University

The IEEE Frontiers '99

Conference provides a major forum for exploring technical issues that are driving the outer boundaries of effective high performance computing. This series of symposia is one of the principal meetings for presenting new and original research results extending the threshold of computational capability through advances in hardware, software, methods, and technology.

The spectrum of fields addressed by the Frontiers conferences includes applications and algorithms, system software and lan-guages, component technologies, and system architectures. A central theme of Frontiers '99 is research related to the exploitation of massive parallelism, and any aspects of the design, analysis, development, and/or use of massively parallel computers. The realms of computing considered will include general purpose, domain specific, and special purpose systems and techniques.

Topics illustrating the spectrum of results ranging from those with near-term practical value to those having long-term implications will be addressed. This dynamic forum provides a stimulating and exciting environment for scientists, engineers, industry representa-tives, and government policy planners to present ideas, findings, product capabilities, and future directions through a series of sessions, panels, and workshops. The conference sessions will be held Tuesday through Thursday; the Workshops will be conducted Sunday afternoon and all day Monday.

- Parallel applications and algorithms, mapping of applications to massively parallel systems, novel algorithmic approaches to problems that are large or irregular nature.
- Very high performance system architectures: MIMD, SIMD, systolic, dataflow, teraflop, petaop systems, intelligent RAM, general-purpose, special-purpose and domain-specific systems, and latency management techniques.
- Resource management for massively parallel computing, languages for highly parallel systems and meta-computing, system software and tools for high per-formance computing, scalable I/O and mass storage.
- Evaluation of system/application scaling and performance.
- Alternative device technologies: optics, superconductors, advanced semiconductors, quantum devices, DNA computing.

14:30-15:00 Optimization of a Parallel Pseudospectral MHD Code
Anshu Dubey
University of Chicago
Thomas Clune
Silicon Graphics, Inc.

15:00-15:30 Material Science Electronic Structure Calculations on Massively Parallel Systems: An Algorithmic and Computational Challenge
Andrew Canning
Lawrence Berkeley National Laboratory

16:00-17:00 Panel: Perspectives on the Accellerated Strategic Computing Initiative
Lisa Thompson, *Director of Government Affairs, Computing Research Association*; David Nowak, *LLNL ASCI Program Leader*; Lawrence Livermore *National Laboratory*; Marc Snir, *Senior Manager, IBM T.J. Watson Research Center TBD, DOE*

18:00 BANQUET

Thursday 25th February

9:00-10:00 Keynote Talk: William R. Pulleyblank, Director, Mathematical Sciences, IBM T.J. Watson Research Center "Deep Computing"

10:30-12:00 Technical presentations

Session 11: Algorithms

10:30-11:00 Asymptotically Optimal Probabilistic Embedding Algorithms for Supporting Tree Structured Computations in Hypercubes
Keqin Li
State University of New York
John Dorband
NASA Goddard Space Flight Center

11:00-11:30 Token Space Minimization by Simulated Annealing
Rafi Lohev, Israel Gottlieb
Bar-Ilan University

11:30-12:00 New Algorithms for Efficient Mining of Association Rules
Hong Shen, Li Shen, Lin Cheng
Griffith University

Session 12: Java for High-Performance Computing

10:30-11:00 Java as a Language for High-Performance Computing
Geoffrey Fox
Syracuse University

11:00-11:30 Java for Numerically Intensive Computing: from Flops to Gigaflops
Marc Snir
IBM

11:30-12:00 Java Numerics: Performance and Portability Issues
Roldan Pozo
NIST

13:30-15:00 Technical Presentations

Session 13 Architecture

13:30-14:00 Superconducting Processors for HTMT: Issues and Challenges
Kevin Theobald
University of Delaware
Guang Gao
McGill University
Thomas Sterling
California Institute of Technology

14:00-14:30 The Preliminary Evaluation of MBP-light with Two Protocol Policies for a Massively Parallel Processor JUMP-1
Hiroaki Inoue, Hideharu Amano, Ken-ichiro Anjo
Keio University
Junji Yamamoto
Real World Computing Partnership
Jun Tanabe, Masaki Wakabayashi
Keio University
Mitsuru Sato
Fujitsu Laboratories
Kei Hiraki
The University of Tokyo

14:30-15:00 Analysis of 100Mb/s Ethernet for the Whitney Commodity Computing Cluster
Samuel Fineberg
Compaq - Tandem Labs
Kevin Pedretti
University of Iowa, Department of ECE

Session 14 Communications

13:30-14:00 Fast Parallel Selection on the Linear Array with Reconfigurable Pipelined Bus System
Yi Pan
University of Dayton
Yijie Han
Electronic Data Systems, Inc.
Hong Shen
Griffith University

14:00-14:30 The Priority Broadcast Scheme for Dynamic Broadcast in Hypercubes and Related Networks
Chi-Hsiang Yeh, Manos Varvarigos, Hua Lee
University of California, Santa Barbara

14:30-15:00 Parallel Algorithms on the Rotation-Exchange Network - A Trivalent Variant of the Star Graph
Chi-Hsiang Yeh, Manos Varvarigos
University of California, Santa Barbara

Appendix D

Advances in Digital Libraries (ADL)

May 19 - 20, 1999

ADL' 99 Exhibit Announcement

Exhibits on the themes Digital Earth and Digital Sky will be shown May 19-20 in Baltimore as part of Advances in Digital Libraries (ADL '99), the premier digital library conference (http://cimic.rutgers.edu/~adl/).

Conference sponsors include NASA/CESDIS and IEEE Computer Society. Admission to the exhibit sessions is free to NASA employees and contractors. PLEASE PRE-REGISTER AT http://cimic.rutgers.edu/~adl/

The Exhibits include commercial exhibits from digital library hardware and technology vendors such as Oracle, Sun, SGI, IBM, ERDAS (a geographic imaging solutions company) and KTI (developing large information based applications). Research projects include the United States Geological Survey; TerraVision, an interactive terrain visualization system from SRI; Profiles in Science and the Visible Human from the NIH/NLM; The National Engineering Education Delivery System (NEEDS) from UC Berkeley; The Informedia Digital Video Library from Carnegie Mellon University; The Art Museum Image Consortium; Digital Meadowlands: An Environmental Digital Library from CIMIC, Rutgers University; The Global Legal Information Network from the Library of Congress.

ADL'99 will be held at the Renaissance Harborplace Hotel, 202 East Pratt Street, Baltimore, MD 21202. Exhibits are May 19 and 20, from 10AM-5PM (closed 1-2PM).

Free registration for the Exhibits Session as well as additional information on the ADL'99 Conference Program is available at http://cimic.rutgers.edu/~adl/

http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html

Appendix E

CESDIS Technical Reports

See the CESDIS Website for a complete set of abstracts

http://cesdis.gsfc.nasa.gov/techreports.html

Nabil Adam, Rutgers University

TR-97-190	**Electronic Commerce and Digital Libraries: Towards a Digital Agora**	**Nabil Adam, Yelena Yesha**	**January 1997**

Electronic commerce (EC) and digital libraries (DL) are two increasingly important areas of computer and information sciences with different user requirements but similar infrastructure requirements. In exploring strategic directions, we examine both requirements of the global information infrastructure that are necessary prerequisite for EC and DL [2], and specific requirements of EC and DL within the global infrastructure.

Both EC and DL are concerned with systems that support the creation of information sources and with the movement of information across global networks. EC supports effective and efficient business interactions and transactions that take place on behalf of consumers, sellers, intermediaries, and producers, while DL supports effective and efficient interaction among knowledge seekers. A digital library may require the transactional aspects of EC to manage the purchasing and distribution of its content while a digital library can be used as a resource in electronic commerce to manage products, services, providers and consumers. EC and DI share a common infrastructure in the networking, security, searching and advertising, negotiating and matchmaking, contracting and ordering, billing, payment, production, distribution, accounting, and customer service mechanisms that support such distributed information systems [31].

In a generic EC/DL model, providers (information providers, merchants, retailers, wholesalers) make multimedia objects available to consumers (customers, information seekers, users) in exchange for payment. An EC/DL system itself is characterized as a collection of distributed autonomous sites (servers) that work together to give the consumer the appearance of a single cohesive collection. Each site may store a large number of multimedia objects (documents, images, video, audio, software, structured data). This content may be stored in a variety of formats and on a variety of media such as disk, tape or CD-ROM and typically originates from a variety of providers who may wish to control its use (retrieval or modification) or to add value. Consumers are assumed to have a wide variety of domain expertise and computer proficiency which must be taken into account by designers of EC/DL systems.

Section 2 examines EC and DL research requirements in six key subareas, which section 3 provides case studies that describe three electronic commerce research projects (USC-ISI, CommerceNet, First Virtual) and six digital libraries projects sponsored by an NSF/ARPA/NASA initiatives.

TR-97-194	**Globalizing Business, Education, Culture Through the Internet**	**Nabil Adam, Baruch Awerbuch, Jacob Slonim, Peter Wegner, Yelena Yesha**	**February 1997**

Globalization occurs at both the national and international levels. Infrastructure is initially developed and regulated at the national level, since most utilization of the telecommunication infrastructure is within rather than among nations. Many of the technical and social questions arising at the national level are relevant to international globalization, while some issues such as interoperability among heterogeneous multilingual components occur primarily at the international level.

The technology of globalization is being driven by commercial incentives for improving the efficiency of business enterprises as well as societal concerns with improving the quality of life. We examine electronic commerce to illustrate business enterprises and education to illustrate the impact of globalization on the quality of life.

Underlying globalization is a set of technologies for human-computer interaction, finding and filtering information, security, negotiating and matchmaking, integration and interoperability, and networking. We discuss a few of these technologies.

TR-97-199	**Information Extraction based Multiple-Category Document Classification for the Global Legal Information Network**	**Nabil Adam, Richard D. Holowczak**	**March 1997**

This paper describes a prototype application of an information extraction (IE) based document classification system in the international law domain. IE is used to determine if a set of concepts for a class are present in a document. The syntactic and semantic constraints that must be satisfied to make this determination are derived automatically from a training corpus. A collection of IE systems are arranged in a classification hierarchy and novel documents are guided down the hierarchy based on a subset of the Global Legal Information Network domain.

TR-97-201	**Modeling and Analysis of Workflows Using Petri Nets**	**Nabil Adam, Vijayalakshmi Aturi, Wei-Kuang Huang**	**April 1997**

A workflow system, in its general form, is basically a heterogeneous and distributed information system where the tasks are performed using autonomous systems. Resources, such as databases, labor, etc. are typically required to process these tasks. Prerequisite to the execution of a task is a set of constraints that reflect the applicable business rules and user requirements.

In this paper we present a *Petri Net* (PN) based framework that (1) facilitates specification of workflow applications, (2) serves as a powerful tool for modeling the system under study at a conceptual level, (3) allows for a smooth transition from the conceptual level to a testbed implementation and (4) enables the analysis, simulation and validation of the system under study before proceeding to implementation. Specifically, we consider three categories of task dependencies: control flow, value, and external (temporal). We identify several structural properties of PN and demonstrate their use for conducting the following type of analyses: (1) identify inconsistent dependency specifications among tasks; (2) test for workflow safety, i.e. test whether the workflow terminates in an acceptable state; (3) for a given starting time, test whether it is feasible to execute a workflow with the specified temporal constraints.

Yair Amir, Johns Hopkins University

TR-98-220	**Seamlessly Selecting the Best Copy from Internet-Wide Replicated Web Servers**	**Yair Amir, Alec Peterson, David Shaw**	

The explosion of the web has led to a situation where a majority of the traffic on the Internet is web related. Today, practically all of the popular web sites are served from single locations. This necessitates frequent long distance network transfers of data (potentially repeatedly) which results in a high response time for users, and is wasteful of the available network bandwidth. Moreover, it commonly creates a single point of failure between the web site and its Internet provider. This paper presents a new approach to web replication, where each of the replicas resides in a different part of the network, and the browser is automatically and transparently directed to the "best" server. Implementing this architecture for popular web sites will result in a better response-time and a higher availability of theses sites. Equally important, this architecture will potentially cut down a significant fraction of the traffic on the Internet, freeing bandwidth for other uses.

Dinshaw Balsara, University of Illinois

TR-98-216	**Analysis of the Eigenstructure, of the Chew, Goldberger and Low System of Equations**	**Dinshaw Balsara, Daniel Spicer**	**September 1997**

The Chew, Goldberger and Low (CGL) System of equations applies to several situations in magnetospheric physics. It is based on making a double adiabatic approximation for the thermal pressure. In this paper we derive the eigenvalues and a complete set of left and right eigenvectors for the CGL system. The system admits eight eigenvalues, seven of which have analogues in ideal MHD. An eighth eigenvalue turns out to correspond to a new kind of advected wave. This wave produces magnetic fluctuations but the magnetic pressure is balanced by the corresponding thermal pressure fluctuation produced by the fact that the thermal pressures are anisotropic. This wave corresponds to a linearly degenerate wave. The eigenvectors for the magnetosonic waves become singular in certain limits. These are identified and eigenvector regularization is done where needed. Intuitive insights pertaining to the nature of the waves are developed. This is especially true for the eighth wave. In the regime of validity of the double adiabatic approximation the wave speeds show a strict ordering. This makes the CGL system amenable to numerical solution using upwind schemes. The linear degeneracy of the eighth wave suggests that it might be treated differently in the context of upwind schemes. Several important parallels as well as some important points of difference between the CGL system of equations and ideal MHD equations are pointed out throughout the paper.

TR-98-217	**Maintaining Pressure Positivity in Magnetohydrodynamics Simulations**	**Dinshaw Balsara, Daniel Spicer**	**December 1997**

Higher order Godunov schemes for solving the equations of Magnetohydrodynamics (MHD) have recently become available. Because such schemes update the total energy, the pressure is a derived variable. In several problems in laboratory physics, magnetospheric physics and astrophysics the pressure can be several orders of magnitude smaller than either the kinetic energy or the magnetic energy. Thus small discretization errors in the total energy can produce situations where the gas pressure can become negative. In this paper we design a linearized Riemann solver that works directly on the entropy density equation. We also design switches that allow us to use such a Riemann solver safely in conjunction with a normal Riemann solver for MHD. This allows us to reduce the discretization errors in the evaluation of the pressure variable. As a result we formulate strategies that maintain the positivity of pressure in all circumstances. We also show via test problems that the strategies designed here work.

TR-98-218	**A Staggered Mesh Algorithm Using High Order Godunov Fluxes to Ensure Solenoidal Magnetic Fields in Magnetohydrodynamic Simulations**	**Dinshaw Balsara, Daniel Spicer**	**December 1997**

The equations of Magnetohydrodynamics (MHD) have been formulated as a hyperbolic system of conservation laws. In that form it becomes possible to use higher order Godunov schemes for their solution. This results in a robust and accurate solution strategy. However, the magnetic field also satisfies a constraint that requires its divergence to be zero at all times. This is a property that cannot be guaranteed in the zone centered discretizations that are favored in Godunov schemes without involving a divergence cleaning step. In this paper we present a staggered mesh strategy which directly uses the properly upwinded fluxes that are provided by a Godunov scheme. The process of directly using the upwinded fluxes relies on a duality that exists between the fluxes obtained from a higher order Godunov scheme and the electric fields in a plasma. By exploiting this duality we have been able to construct a higher order Godunov scheme that ensures that the magnetic field remains divergence free up to the computer's round-off error. Several stringent test problems have been devised to show that the scheme works robustly and accurately in all situations. In doing so it is shown that a scheme that involves a collocation of magnetic field variable that is different from the one traditionally favored in the design of higher order Godunov schemes can nevertheless offer the same robust and accurate performance of higher order Godunov schemes provided the properly upwinded fluxes from the Godunov methodology are used in the scheme's construction.

Donald Becker, CESDIS

TR-98-214	**An Assessment of Beowulf-class, Computing for NASA Requirements: Initial Findings from the First NASA Workshop on Beowulf-class Clustered Computing**	**Donald Becker, Thomas Sterling, Mike Warren, Tom Cwik, John Salmon, Bill Nitzberg**	**January 1998**

The Beowulf class of parallel computing machine started as a small research project at NASA Goddard Space Flight Center's Center of Excellence in Space Data and Information Sciences (CESDIS). From that work evolved a new class of scalable machine comprised of mass market common off-the-shelf components (M^2COTS) using a freely available operating system and industry-standard software packages. A Beowulf-class system provides extraordinary benefits in price-performance. Beowulf-class systems are in place and doing real work at several NASA research centers, are supporting NASA-funded academic research, and operating at DOE and NIH. The NASA user community conducted an intense two-day workshop in Pasadena, California on October 22-23, 1997. This first workshop on Beowulf-class systems consisted primarily of technical discussions to establish the scope of opportunities, challenges, current research activities, and directions for NASA computing employing Beowulf-class systems. The technical discussions ranged from application research to programming methodologies. This paper provides an overview of the findings and conclusions of the workshop. The workshop determined that Beowulf-class systems can deliver multi-Gflops performance at unprecedented price-performance but that software environments were not fully functional or robust, especially for larger "dreadnought" scale systems. It is recommended that the Beowulf community engage in an activity to integrate, port, or develop, where appropriate, necessary components of the software infrastructure to fully realize the potential of Beowulf-class computing to meet NASA and other agency computing requirements.

TR-98-219	**Achieving Ten Gflops on PC Clusters: A Case Study**	**Udaya Ranawake, John Dorband, Bruce Fryxell, Daniel Ridge, Erik Hendriks, Donald Becker, Phillip Merkey**	**May 1998**

The Beowulf project is a NASA Initiative to harness the parallelism of PC clusters built from commodity microprocessors and networking hardware and to develop the technology to apply these systems to NASA earth and space science computational needs. In this paper, we describe a case study using an important space science application that achieves more than 10 Gflops on 199 processors of a Beowulf class PC cluster. This represents nearly a ten fold increase in performance for this class of computer systems within one year. We describe the methodologies used to achieve this breakthrough and discuss the results from benchmarking runs that compare the performance of these systems with high end supercomputers such as the Cray T3E and the Convex SPP 2000.

Key words: Beowulf project, PC clusters, benchmarks, performance evaluation, scalability.

Diane Cook, University of Texas at Arlington

TR-98-222	**Parallel Knowledge Discovery from Large Complex Databases**	**Diane Cook, Lawrence Holder**	

Nasa is focusing on grand challenge problems in Earth and space sciences. Within these areas of science, new instrumentation will be providing scientists with unprecedented amounts of unprocessed data. Our goal is to design and implement a system that takes raw data as input and efficiently discovers interesting concepts that can target areas for further investigation and can be used to compress the data. Our approach will provide an intelligent parallel data analysis system.

Tarek El-Ghazawi, George Mason University

TR-97-203	**Wavelet-Based Image Registration on Parallel Computers**	**Tarek El-Ghazawi, Prachya Chalermwat, Jacqueline Le Moigne**	**November 1997**

Digital image registration is very important in many applications, such as medical imagery, robotics, visual inspection, and remotely sensed data processing. In particular, NASA's Mission To Planet Earth (MTPE) program will be producing enormous Earth global change data, reaching hundreds of Gigabytes per day, that are collected from different spacecraft's and different perspectives using many sensors with diverse resolution and characteristics. The analysis of such data requires integration, therefore, accurate registration of these data. Image registration is defined as the process which determines the most accurate relative orientation between two or more images, acquired at the same or different times by different or identical sensors. Registration can also provide the absolute orientation between an image and a map.

Erik Hendriks, CESDIS

TR-98-219	**Achieving Ten Gflops on PC Clusters: A Case Study**	**Udaya Ranawake, John Dorband, Bruce Fryxell, Daniel Ridge, Erik Hendriks, Donald Becker, Phillip Merkey**	**May 1998**

The Beowulf project is a NASA Initiative to harness the parallelism of PC clusters built from commodity microprocessors and networking hardware and to develop the technology to apply these systems to NASA earth and space science computational needs. In this paper, we describe a case study using an important space science application that achieves more than 10 Gflops on 199 processors of a Beowulf class PC cluster. This represents nearly a ten fold increase in performance for this class of computer systems within one year. We describe the methodologies used to achieve this breakthrough and discuss the results from benchmarking runs that compare the performance of these systems with high end supercomputers such as the Cray T3E and the Convex SPP 2000.

Key words: Beowulf project, PC clusters, benchmarks, performance evaluation, scalability.

Jacqueline LeMoigne, CESDIS

TR-97-203	**Wavelet-Based Image Registration on Parallel Computers**	**Tarek El-Ghazawi, Prachya Chalermwat, Jacqueline Le Moigne**	**November 1997**

Digital image registration is very important in many applications, such as medical imagery, robotics, visual inspection, and remotely sensed data processing. In particular, NASA's Mission To Planet Earth (MTPE) program will be producing enormous Earth global change data, reaching hundreds of Gigabytes per day, that are collected from different spacecraft's and different perspectives using many sensors with diverse resolution and characteristics. The analysis of such data requires integration, therefore, accurate registration of these data. Image registration is defined as the process which determines the most accurate relative orientation between two or more images, acquired at the same or different times by different or identical sensors. Registration can also provide the absolute orientation between an image and a map.

TR-97-206	**Proceedings of the Image Registration Workshop**	**Jacqueline Le Moigne**	**November 1997**

Automatic image registration has often been considered as a preliminary step for higher-level processing, such as object recognition or data fusion, but with the unprecedented amounts of data which are being and will continue to be generated by newly developed sensors. The very topic of automatic image registration has become an important research topic. The Image Registration Workshop (IRW '97), which was held at NASA/Goddard Space Flight Center on November 20-21, was one of the first to concentrate on the issue of automatic image registration. These workshop proceedings present a collection of very high quality work which has been grouped into four main areas: (1) theoretical aspects of image registration, (2) applications to satellite imagery, (3) applications to medical imagery, (4) image registration for computer vision research.

TR-97-207	**Satellite Imaging and Sensing**	**Jacqueline Le Moigne, Robert F. Cromp**	**November 1997**

Satellite imaging and sensing is the process by which the electromagnetic energy reflected or emitted from any planetary surface is captured by a sensor located on a spaceborne platform. This article describes the general principles and characteristics related to satellite sensors as well as examples of some typical attributes which can be measured from space. A summary of most of the principal earth remote sensing systems is given, and a few space applications are described. Management and interpretation of data acquired by satellite is a very important issue and this article summarizes some preliminary ideas on how the digital representation is formed and the basic types of data processing necessary before any further interpretation of the data. As a conclusion, the future in satellite imaging and sensing is briefly addressed.

TR-98-221	**An Evaluation of Automatic Image Registration Methods**	**Jacqueline Le Moigne, Wei Xia, James Tilton, Prachya Chalermwat, Tarek El-Ghazawi, Nathan Netanyahu, David Mount, William Campbell**	

The study of global environmental changes involves the comparison, fusion, and integration of multiple types of remotely-sensed data at various temporal, radiometric, and spatial resolutions. Results of this integration may be utilized for global change analysis, as well as for the validation of new instruments or for new data analysis. Furthermore, smaller missions will include many different sensors carried on separate platforms, and the amount to remote sensing data to be combined will increase tremendously. For all of these applications, the first required step is fast and automatic image registration.

As the need for automating registration techniques is being recognized, it becomes necessary to survey all the registration methods which may be applicable to Earth and space science problems and to evaluate their performances on a large variety of existing remote sensing data as well as on simulated data of soon-to-be-flown instruments. In this paper we present the first steps toward an exhaustive quantitative evaluation: four automatic image registration algorithms are described and results of their evaluation are presented on three different datasets. The four algorithms are based on gray levels, edge features or wavelet features and compute translation, similarity or rigid transformations. Results show that the four selected methods are within 2 pixel accuracy, and that a tradeoff must be achieved between computation time and accuracy of the computed deformation.

Richard Lyon, University of Maryland Baltimore County

TR-97-196	**Hubble Space Telescope Faint Object Camera Calculated Point-Spread Functions**	**Rick Lyon, Jan M. Hollis, John E. Dorband**	**March 1997**

A set of observed noisy Hubble Space Technology Faint Camera point-spread functions used to recover the combined Hubble and Faint Object Camera wave-front error. The low-spatial-frequency wave-front error is parameterized in terms of a set of 32 annular Zernike polynomials. The midlevel and higher spatial frequencies are parameterized in terms of set of 891 polar-Fourier polynomials. The parameterized wave-front error is used to generate accurate calculated point-spread functions, both pre- and post-COSTAR (corrective optics space telescope axial replacement), suitable for image restoration at arbitrary wavelengths. We describe the phase-retrieval-based recovery process and the phase parameterization. Resultant calculated precorrection and postcorrection point-spread functions are shown along with an estimate of both pre- and post-COSTAR spherical aberration.

TR-97-197	**Motion of the Ultraviolet R Aquarii Jet**	**Rick Lyon, Jan M. Hollis, John E. Dorband, W.A. Feibelman**	**January 1997**

We present evidence for subarcsecond changes in the ultraviolet (~2550 Å) morphology of the inner 5 arcseconds of the R Aqr jet over a 2 yr. period. These data were taken with the *Hubble Space Telescope* (HST) Faint Object Camera (FOC) when the primary mirror flow was still affecting observations. Images of the R Aqr stellar jet were successfully restored to the original design resolution by completely characterizing the telescope-camera point spread function (PSF) with the aid of phase-retrieval techniques. Thus, a noise-free PSF was employed in the final restorations which utilized the maximum entropy method (MEM). We also present recent imagery obtained with the HST/FOC system after the COSTAR correction mission that provides confirmation of the validity of our restoration methodology. The restored results clearly show that the jet is flowing along the northeast (NE)-southwest (SW) axis with a prominent helical-like structure evident on the stronger NE side of the jet. Transverse velocities increase with increasing distance from the central source, providing a velocity range of 36-235 km s^{-1}. From an analysis of proper motions of the two major ultraviolet jet components, we detect an ~40.2 yr. event separation of this apparent enhanced material ejection occurring probably at periastron which is consistent with the suspected ~44 yr. binary period; this same analysis shows that the jet is undergoing magnetic effects. The restoration computations and the algorithms employed demonstrate that mining of flawed HST data can be scientifically worthwhile.

TR-97-198	**A Maximum Entropy Method with a Priori Maximum Likelihood**	**Rick Lyon, Jan M. Hollis, John E. Dorband**	**April 1997**

Implementations of the maximum entropy method for data reconstruction have almost universally used the approach of maximizing the statistic S - $\lambda\chi^2$ where S is the Shannon entropy of the reconstructed distribution and χ^2 is the usual statistical measure associated with agreement between certain properties of the reconstructed distribution and the data. We develop here an alternative approach which maximizes the entropy subject to the set of constraints the χ^2 be at a minimum with respect to the reconstructed distribution. This in turn modifies the fitting statistics to be $S - \lambda \bullet \nabla\chi^2$ where λ is now a vector. This new method provided a unique solution to both the well-posed and ill-posed problem, provides a natural convergence criterion which has previously been lacking in other implementations of maximum entropy, and provides the most conservative (least informative) data reconstruction result consistent with both maximum entropy and maximum likelihood methods, thereby mitigating against over-interpretation of reconstruction results. A spectroscopic example is shown as a demonstration.

Daniel Menascé, George Mason University

TR-97-188	**Pythia and Pythia/ WK: Tools for the Performance Analysis of Mass Storage Systems**	**Odysseas I. Pentakalos, Daniel A. Menascé, Yelena Yesha**	**January 1997**

The constant growth on the demands imposed on hierarchical mass storage systems creates a need for frequent reconfiguration and upgrading to ensure that the response times and other performance metrics are within the desired service levels. This paper describes the design and operation of two tools, Pythia and Pythia/WK, that assist system managers and integrators in making cost-effective procurement decisions. Pythia automatically builds and solves an analytic model of a mass storage system based on a graphical description of the architecture of the system and on a description of the workload imposed the system. The use of a modeling wizard to perform this conversion unique among analytic performance tools. Pythia/WK uses clustering algorithms to characterize the workload from the log files of the mass storage system. The resulting workload characterization is used as input to Pythia.

TR-97-202	**Pythia: A Performance Analyzer of Hierarchical Mass Storage Systems**	**Odysseas Pentakalos, Daniel Menascé, Yelena Yesha**	**July 1997**

Hierarchical mass storage systems are becoming more complex each day and there are many possible ways of configuring them. The options range from the type an number of devices to be used to their connectivity. An extensible object-oriented performance analyzer, called Pythia, was designed and implemented to allow users to easily investigate the most cost-effective configurations for a given workload. One of the most important reasons to build such a tool is to provide a simple way through which queuing analytic models can be used for performance prediction and system sizing of mass storage systems. The tool incorporated a modeling wizard component that is capable of automatically building a queuing network model from a mass storage system representation defined through a graphic editor. Thus, the user of the tool does not need to know queuing network modeling techniques to use it.

Phillip Merkey, CESDIS

TR-98-219	**Achieving Ten Gflops on PAC Clusters: A Case Study**	**Udaya Ranawake, John Dorband, Bruce Fryxell, Daniel Ridge, Erik Hendriks, Donald Becker, Phillip Merkey**	**May 1998**

The Beowulf project is a NASA Initiative to harness the parallelism of PC clusters built from commodity microprocessors and networking hardware and to develop the technology to apply these systems to NASA earth and space science computational needs. In this paper, we describe a case study using an important space science application that achieves more than 10 Gflops on 199 processors of a Beowulf class PC cluster. This represents nearly a ten fold increase in performance for this class of computer systems within one year. We describe the methodologies used to achieve this breakthrough and discuss the results from benchmarking runs that compare the performance of these systems with high end supercomputers such as the Cray T3E and the Convex SPP 2000.

Key words: Beowulf project, PC clusters, benchmarks, performance evaluation, scalability.

Udaya Ranawake, University of Maryland Baltimore County

TR-98-219	**Achieving Ten Gflops on PC Clusters: A Case Study**	**Udaya Ranawake, John Dorband, Bruce Fryxell, Daniel Ridge, Erik Hendriks, Donald Becker, Phillip Merkey**	**May 1998**

The Beowulf project is a NASA Initiative to harness the parallelism of PC clusters built from commodity microprocessors and networking hardware and to develop the technology to apply these systems to NASA earth and space science computational needs. In this paper, we describe a case study using an important space science application that achieves more than 10 Gflops on 199 processors of a Beowulf class PC cluster. This represents nearly a ten fold increase in performance for this class of computer systems within one year. We describe the methodologies used to achieve this breakthrough and discuss the results from benchmarking runs that compare the performance of these systems with high end supercomputers such as the Cray T3E and the Convex SPP 2000.

Key words: Beowulf project, PC clusters, benchmarks, performance evaluation, scalability.

Daniel Ridge, University of Maryland College Park

TR-98-219	**Achieving Ten Gflops on PC Clusters: A Case Study**	**Udaya Ranawake, John Dorband, Bruce Fryxell, Daniel Ridge, Erik Hendriks, Donald Becker, Phillip Merkey**	**May 1998**

The Beowulf project is a NASA Initiative to harness the parallelism of PC clusters built from commodity microprocessors and networking hardware and to develop the technology to apply these systems to NASA earth and space science computational needs. In this paper, we describe a case study using an important space science application that achieves more than 10 Gflops on 199 processors of a Beowulf class PC cluster. This represents nearly a ten fold increase in performance for this class of computer systems within one year. We describe the methodologies used to achieve this breakthrough and discuss the results from benchmarking runs that compare the performance of these systems with high end supercomputers such as the Cray T3E and the Convex SPP 2000.

Key words: Beowulf project, PC clusters, benchmarks, performance evaluation, scalability.

Yelena Yesha, CESDIS and University of Maryland Baltimore County

TR-97-188	**Pythia and Pythia/ WK: Tools for the Performance Analysis of Mass Storage Systems**	**Odysseas I. Pentakalos, Daniel A. Menascé, Yelena Yesha**	**January 1997**

The constant growth on the demands imposed on hierarchical mass storage systems creates a need for frequent reconfiguration and upgrading to ensure that the response times and other performance metrics are within the desired service levels. This paper describes the design and operation of two tools, Pythia and Pythia/WK, that assist system managers and integrators in making cost-effective procurement decisions. Pythia automatically builds and solves an analytic model of a mass storage system based on a graphical description of the architecture of the system and on a description of the workload imposed the system. The use of a modeling wizard to perform this conversion unique among analytic performance tools. Pythia/WK uses clustering algorithms to characterize the workload from the log files of the mass storage system. The resulting workload characterization is used as input to Pythia.

TR-97-190	**Electronic Commerce and Digital Libraries: Towards a Digital Agora**	**Nabil Adam, Yelena Yesha**	**January 1997**

Electronic commerce (EC) and digital libraries (DL) are two increasingly important areas of computer and information sciences with different user requirements but similar infrastructure requirements. In exploring strategic directions, we examine both requirements of the global information infrastructure that are necessary prerequisite for EC and DL [2], and specific requirements of EC and DL within the global infrastructure.

Both EC and DL are concerned with systems that support the creation of information sources and with the movement of information across global networks. EC supports effective and efficient business interactions and transactions that take place on behalf of consumers, sellers, intermediaries, and producers, while DL supports effective and efficient interaction among knowledge seekers. A digital library may require the transactional aspects of EC to manage the purchasing and distribution of its content while a digital library can be used as a resource in electronic commerce to manage products, services, providers and consumers. EC and DI share a common infrastructure in the networking, security, searching and advertising, negotiating and matchmaking, contracting and ordering, billing, payment, production, distribution, accounting, and customer service mechanisms that support such distributed information systems [31].

In a generic EC/DL model, providers (information providers, merchants, retailers, wholesalers) make multimedia objects available to consumers (customers, information seekers, users) in exchange for payment. An EC/DL system itself is characterized as a collection of distributed autonomous sites (servers) that work together to give the consumer the appearance of a single cohesive collection. Each site may store a large number of multimedia objects (documents, images, video, audio, software, structured data). This content may be stored in a variety of formats and on a variety of media such as disk, tape or CD-ROM and typically originates from a variety of providers who may wish to control its use (retrieval or modification) or to add value. Consumers are assumed to have a wide variety of domain expertise and computer proficiency which must be taken into account by designers of EC/DL systems.

Section 2 examines EC and DL research requirements in six key subareas, which section 3 provides case studies that describe three electronic commerce research projects (USC-ISI, CommerceNet, First Virtual) and six digital libraries projects sponsored by an NSF/ARPA/NASA initiatives.

TR-97-194	**Globalizing Business, Education, Culture Through the Internet**	**Nabil Adam, Baruch Awerbuch, Jacob Slonim, Peter Wegner, Yelena Yesha**	**February 1997**

Globalization occurs at both the national and international levels. Infrastructure is initially developed and regulated at the national level, since most utilization of the telecommunication infrastructure is within rather than among nations. Many of the technical and social questions arising at the national level are relevant to international globalization, while some issues such as interoperability among heterogeneous multilingual components occur primarily at the international level.

The technology of globalization is being driven by commercial incentives for improving the efficiency of business enterprises as well as societal concerns with improving the quality of life. We examine electronic commerce to illustrate business enterprises and education to illustrate the impact of globalization on the quality of life.

Underlying globalization is a set of technologies for human-computer interaction, finding and filtering information, security, negotiating and matchmaking, integration and interoperability, and networking. We discuss a few of these technologies.

TR-97-202	**Pythia: A Performance Analyzer of Hierarchical Mass Storage Systems**	**Odysseas Pentakalos, Daniel Menascé, Yelena Yesha**	**July 1997**

Hierarchical mass storage systems are becoming more complex each day and there are many possible ways of configuring them. The options range from the type an number of devices to be used to their connectivity. An extensible object-oriented performance analyzer, called Pythia, was designed and implemented to allow users to easily investigate the most cost-effective configurations for a given workload. One of the most important reasons to build such a tool is to provide a simple way through which queuing analytic models can be used for performance prediction and system sizing of mass storage systems. The tool incorporated a modeling wizard component that is capable of automatically building a queuing network model from a mass storage system representation defined through a graphic editor. Thus, the user of the tool does not need to know queuing network modeling techniques to use it.

CESDIS

CENTER OF EXCELLENCE IN SPACE DATA AND INFORMATION SCIENCES
CODE 930.5
NASA GODDARD SPACE FLIGHT CENTER
GREENBELT, MD 20771

TECHNICAL REPORT SERIES ORDER FORM

301-286-4403 **Internet: cas@cesdis1.gsfc.nasa.gov**

Number of Copies Requested	Report Number	Title
________	TR-90-01	Analyzing a CSMA/CD Protocol through a Systems of Communicating Machines Specification (Raymond E. Miller)
________	TR-90-02	Altruistic Locking (Kenneth Salem)
________	TR-90-03	Modeling the Logical Structure of Flexible Manufacturing Systems with Petri-Nets (P. David Stotts)
________	TR-90-04	On the Bit-Complexity of Discrete Solutions of PDEs: Compact Multigrid (John Reif)
________	TR-90-05	Rules and Principles of Scientific Data Visualization (Hikmet Senay)
________	TR-90-06	Changes in Connectivity in Active Contour Models (Ramin Samadani)
________	TR-90-07	Designing C++ Libraries (James M. Coggins)
________	TR-90-08	Stabilization and Pseudo-Stabilization (Raymond E. Miller)
________	TR-90-09	Coordinating Multi-Transaction Activities (Kenneth Salem)
________	TR-90-10	Bounding Procedure Execution Times in a Synchronous (P. David Stotts)
________	TR-90-11	VISTA: Visualization Tool Assistant for Viewing Scientific Data (Hikmet Senay)
________	TR-90-12	Model-Driven Image Analysis to Augment Databases (Ramin Samadani)
________	TR-90-13	Interfacing Image Processing and Computer Graphics Systems Using an Artificial Visual System (James M. Coggins)

________	TR-90-14	Protocol Verification: The First Ten Years, The Next Ten Years; Some Personal Observations (Raymond E. Miller)
________	TR-90-15	Coverability Graphs for a Class of Synchronously Executed Unbounded Petri Net (P. David Stotts)
________	TR-90-16	Compositional Analysis and Synthesis of Scientific Data Visualization Techniques (Hikmet Senay)
________	TR-90-17	Evaluation of an Elastic Curve Technique for Automatically Finding the Auroral Oval from Satellite Images (Ramin Samadani)
________	TR-90-18	Anticipated Methodologies in Computer Vision (James M. Coggins)
________	TR-90-19	Synthesizing a Protocol Converter from Executable Protocol Traces (Raymond E. Miller)
________	TR-90-20	YTRACC: An Interactive Debugger for YACC Grammars (David P. Stotts)
________	TR-90-21	Finding Curvilinear Features in Speckled Images (Ramin Samadani)
________	TR-90-22	Multiscale Geometric Image Descriptions for Interactive Object Definition (James M. Coggins)
________	TR-90-23	Testing Protocol Implementations Based on a Formal Specification (Raymond E. Miller)
________	TR-90-24	A Mills-Style Iteration Theorem for Nondeterministic Concurrent Program (P. David Stotts)
________	TR-90-25	A Computer Vision System for Automatically Finding the Auroral Oval from Satellite Images (Ramin Samadani)
________	TR-90-26	Multiscale Vector Fields for Image Pattern Recognition (James M. Coggins)
________	TR-90-28	Generalizing Hypertext (P. David Stotts)
________	TR-90-29	Evaluation of an Elastic Curve Technique for Automatically Finding the Auroral Oval from Satellite Images (Ramin Samadani)
________	TR-90-30	Interactive Object Definition in Medical Images Using Multiscale, Geometric Image Descriptions (James M. Coggins)
________	TR-90-31	Two New Approaches to Conformance Testing of Communication Protocols (Raymond E. Miller)
________	TR-90-32	Increasing the Power of Hypertext Search with Relational Queries (P. David Stotts)
________	TR-90-33	A Multiscale Description of Image Structure for Segmentation of Biomedical Images (James M. Coggins)
________	TR-90-34	Temporal Hyperprogramming (P. David Stotts)

________	TR-90-35	Biomedical Image Segmentation Using Multiscale Orientation Fields (James M. Coggins)
________	TR-90-36	Programmable Browsing Semantics in Trellis (P. David Stotts)
________	TR-90-37	Image Structure Analysis Supporting Interactive Object Definition (James M. Coggins)
________	TR-90-38	Separating Hypertext Content from Structure in Trellis (P. David Stotts)
________	TR-90-39	Hierarchy, Composition, Scripting Languages, and Translators for Structured Hypertext (P. David Stotts)
________	TR-90-40	Browsing Parallel Process Networks (P. David Stotts)
________	TR-90-41	aTrellis: A System for Writing and Browsing Petri-Net-Based Hypertext (P. David Stotts)
________	TR-91-42	Generating Test Sequences with Guaranteed Fault Coverage for Conformance Testing of Communication Protocols (Raymond E. Miller)
________	TR-91-43	Specification and Analysis of a Data Transfer Protocol Using Systems of Communicating Machines (Raymond E. Miller)
________	TR-91-44	An Exact Algorithm for Kinodynamic Planning in the Plane (John Reif)
________	TR-91-45	Place/Transition Nets with Debit Arcs (P. David Stotts, Parke Godfrey)
________	TR-91-46	Adaptive Prefetching for Disk Buffers (Kenneth Salem)
________	TR-91-48	Adaptive Control of Parameters for Active Contour Models (Ramin Samadani) Visual System (James M. Coggins)
________	TR-91-50	Structured Dynamic Behavior in Hypertext (David Stotts)
________	TR-91-51	BLITZEN: A Highly Integrated Massively Parallel Machine (John Reif)
________	TR-91-52	Efficient Parallel Algorithms for Optical Computing with the DFT Primitive (John Reif)
________	TR-91-53	This Technical Report has been superceded by TR-92-87 A Minimization-Pruning Algorithm for Finding Elliptical Boundaries in Images with Non-Constant Background and with Missing Data (Ramin Samadani)
________	TR-91-54	A Functional Meta-Structure for Hypertext Models and Systems (P. David Stotts)
________	TR-91-55	An Optimal Parallel Algorithm for Graph Planarity (John Reif)
________	TR-91-56	A Randomized EREW Parallel Algorithm for Finding Connected Components in a Graph (Hillel Gazit and John Reif)
________	TR-91-57	Study of Six Linear Least Square Fits (Eric Feigelson)
________	TR-91-58	Fast Computations of Vector Quantization Algorithms (John Reif)

________	TR-91-59	Probabilistic Diagnosis of Hot Spots (Kenneth Salem)
________	TR-91-60	Multi-Media Interaction with Virtual Worlds (Hikmet Senay)
________	TR-91-61	Image Compression Methods with Distortion Controlled Capabilities (John Reif)
________	TR-91-62	Management of Partially-Safe Buffers (Kenneth Salem)
________	TR-91-63	Non-Deterministic Queue Operations (Kenneth Salem)
________	TR-91-64	Dynamic Adaptation of Hypertext Structure (P. David Stotts)
________	TR-91-65	Scientific Data Visualization Software: Trends and Directions (James Foley)
________	TR-91-66	Planar Separators and the Euclidean Norm (Hillel Gazit)
________	TR-91-67	A Deterministic Parallel Algorithm for Planar Graphs Isomorphism (Hillel Gazit)
________	TR-91-68	A Deterministic Parallel Algorithm for Finding a Separator in Planar Graphs (Hillel Gazit)
________	TR-91-69	An Algorithm for Finding a Seperator in Planar Graphs (Hillel Gazit)
________	TR-91-70	Optimal EREW Parallel Algorithms for Connectivity Ear Decomposition and st- Numbering of Planar Graphs (Hillel Gazit)
________	TR-91-71	An Optimal Randomized Parallel Algorithm for Finding Connecting Components in a Graph (Hillel Gazit)
________	TR-91-72	Modified Version of Generating Minimal Length Test Sequences for Conformance Testing of Communication Protocols (Raymond Miller)
________	TR-91-73	Adaptive Image Segmentation Applied to Extracting the Auroral Oval from Satellite Images (Ramin Samadani)
________	TR-91-74	Parallel Programming on the Silicon Graphics Workstation Using the Multiprocessing Library (Cynthia Starr)
________	TR-91-75	Placing Replicated Data to Reduce Seek Delays (Kenneth Salem)
________	TR-92-76	CESDIS Annual Report; Year 3
________	TR-92-77	Generating Test Sequences with Guaranteed Fault Coverage for Conformance Testing of Communication Protocols (Raymond E. Miller)
________	TR-92-78	On the Generation of Minimal Length Conformance Tests for Communication Protocols (Raymond E. Miller)
________	TR-92-79	A Knowledge Based System for Scientific Data Visualization (Hikmet Senay)

________	TR-92-80	On Generating Test Sequences for Combined Control and Data Flow for Conformance Testing of Communication Protocols (Raymond E. Miller)
________	TR-92-81	MR-CDF: Managing Multi-Resolution Scientific Data (Kenneth Salem)
________	TR-92-82	Adaptive Block Rearrangement (Kenneth Salem)
________	TR-92-83	A Markov Field/Accumulator Sampler Approach to the Atmospheric Temperature Inversion Problem (Noah Friedland)
________	TR-92-84	Adaptive Snakes: Control of Damping and Material Parameters (Ramin Samadani)
________	TR-92-85	Faults, Errors and Convergence in Conformance Testing of Communication Protocols (Raymond E. Miller, Sanjoy Paul))
________	TR-92-86	Research Issues for Communication Protocols (Raymond E. Miller)
________	TR-92-87	This Technical Report Supercedes TR-91-53 A Minimization-Pruning Algorithm for Finding Elliptical Boundaries in Images with Non-Constant Background and with Missing Data (Ramin Samadani)
________	TR-92-88	Summary Report of the CESDIS Workshop on Scientific Database Management (Kenneth Salem)
________	TR-92-89	Structural Analysis of a Protocol Specification and Generation of a Maximal Fault Coverage Conformance Test Sequence (Raymond E. Miller)
________	TR-92-90	Kernel-Control Parallel Versus Data Parallel: A Technical Comparison (Terrence Pratt)
________	TR-92-91	Efficient Synchronization with Minimal Hardware Support (James H. Anderson)
________	TR-92-92	A Fine-Grained Solution to the Mutual Exclusion Problem (James H. Anderson)
________	TR-92-93	On the Granularity of Conditional Operations (James H. Anderson, Mohamed G. Gouda)
________	TR-92-94	CESDIS Annual Report; Year 4
________	TR-93-95	Image Analysis by Integration of Disparate Information (Jacqueline Le Moigne)
________	TR-93-96	A Virtual Machine for High Performance Image Processing (Douglas Smith)
________	TR-93-97	Generating Maximal Fault Coverage Conformance Test Sequences of Reduced Length for Communication Protocols (Raymond E. Miller)
________	TR-93-98	Bounding the Performance of FDDI (Raymond E. Miller)
________	TR-93-99	Report on the Workshop on Data and Image Compression Needs and Uses in Scientific Community (Stephen R. Tate)

________	TR-93-100	Fine Grain Dataflow Computation without Tokens for Balanced Execution (Thomas Sterling)
________	TR-93-101	Implementing Extended Transaction Models Using Transaction Groups (Kenneth Salem)
________	TR-93-102	Adaptive Block Rearrangement Under UNIX (Kenneth Salem)
________	TR-93-103	The Realities of High Performance Computing and Dataflow's Role in It: Lessons from the NASA HPCC Program (Thomas Sterling)
________	TR-93-104	Summary Report of the CESDIS Seminar Series on Earth Remote Sensing (Jacqueline Le Moigne)
________	TR-93-105	Space-Efficient Hot Spot Estimation (Kenneth Salem)
________	TR-93-106	DQDB Performance and Fairness as Related to Transmission Capacity (Raymond E. Miller)
________	TR-93-107	Deadlock Detection for Cyclic Protocols Using Generalized Fair Reachability Analysis (Raymond E. Miller)
________	TR-94-108	Summary Report of the CESDIS Seminar Series on Future Earth Remote Sensing Missions (Jacqueline Le Moigne)
________	TR-94-109	Generalized Fair Reachability Analysis for Cyclic Protocols: Part 1 (Raymond E. Miller)
________	TR-94-110	CESDIS Annual Report; Year 5
________	TR-94-111	This Technical Report has been superceded by TR-94-129. I/O Performance of the MasPar MP-1 Testbed (Tarek A. El-Ghazawi)
________	TR-94-112	Parallel Registration of Multi-Sensor Remotely Sensed Imagery Using Wavelet Coefficients (Jacqueline Le Moigne)
________	TR-94-113	Paradise - A Parallel Geographic Information System (David De Witt)
________	TR-94-114	Computer Assisted Analysis of Auroral Images Obtained from High Altitude Polar Satellites (Ramin Samadani)
________	TR-94-115	2Q: A Low Overhead High Performance Buffer Management Replacement Algorithm (Theodore Johnson)
________	TR-94-116	Sensitivity Analysis of Frequency Counting (Theodore Johnson)
________	TR-94-117	Client-Server Paradise (David De Witt)
________	TR-94-118	Performance Characteristics of a 100 MegaByte/second Disk Array (Matthew T. O'Keefe)
________	TR-94-119	Compiler and Runtime Support for Out-of-Core HPF Programs (Alok Choudhary)

________	TR-94-120	Use of Subband Decomposition for Management of Scientific Image Databases (Kathleen G. Perez-Lopez)
________	TR-94-121	Client-Server Paradise (David DeWitt)
________	TR-94-122	Multi-Resolution Wavelet Decomposition on the MasPar Massively Parallel System (Jacqueline LeMoigne and Tarek A. El-Ghazawi)
________	TR-94-123	An Initial Evaluation of the Convex SPP-1000 for Earth and Space Science Applications (Thomas Sterling, Phillip Merkey)
________	TR-94-124	Runtime Support for Parallel I/O in PASSION (Alok Choudhary)
________	TR-94-125	The Performance Impact of Data Placement for Wavelet Decomposition of Two Dimensional Image Data on SIMD Machines (Jacqueline LeMoigne and Tarek El-Ghazawi)
________	TR-94-126	Planet Photo-Topography Using Shading and Stereo (Charles XiaoJian Yan)
________	TR-94-127	Highly Scalable Data Balanced Distributed B-trees (Theodore Johnson)
________	TR-94-128	Index Replication in a Distributed B-tree (Theodore Johnson)
________	TR-94-129	Characteristics of the MasPar Parallel I/O System (Tarek El-Ghazawi)
________	TR-94-130	PASSION: Parallel and Scalable Software for Input-Output (Alok Choudhary)
________	TR-94-131	Development of a Data Reduction Expert Assistant (Glenn Miller)
________	TR-94-132	Multivariate Statistical Analysis Software Technologies for Astrophysical Research Involving Large Data Sets (S. G. Djorgovski)
________	TR-94-133	The Grid Analysis and Display System (GrADS) (James Kinter)
________	TR-94-134	An Interactive Environment for the Analysis of Large Earth Observation and Model Data Sets (Kenneth Bowman and Robert Wilhelmson)
________	TR-94-135	A Distributed Analysis and Visualization System for Model and Observational Data (Robert Wilhelmson)
________	TR-94-136	VIEWCACHE: An Incremental Database Access Method for Autonomous Interoperable Databases (Nicholas Roussopoulos)
________	TR-94-137	Topography from Shading and Stereo (Berthold Horn)
________	TR-94-138	Experimenter's Laboratory for Visualized Interactive Science (Elaine Hansen)
________	TR-94-139	A Land-Surface Testbed for EOSDIS (William Emery)
________	TR-94-140	High Performance Compression of Science Data (James Storer)

________	TR-94-141	SAVS: A Space and Atmospheric Visualization Science System (Edward P. Szuszcwicz)
________	TR-94-142	Interactive Interface for NCAR Graphics (Bill Buzbee)
________	TR-94-143	McIDAS-eXplorer: A Tool for Analysis of Planetary Data (Sanjay Limaye)
________	TR-94-144	Software-based Fault Tolerance (Jonathan Bright)
________	TR-95-145	AstroNet: A Tool Set for Simultaneous, Multi-Site Observations of Astronomical Objects (Supriya Chakrabarti)
________	TR-95-146	Refining Image Segmentation by Integration of Edge and Region Data (Jacqueline Le Moigne and James Tilton)
________	TR-95-147	An Approximate Performance Model of a Unitree Mass Storage System (Odysseas I. Pentakalos, Daniel A. Menasce, Milt Halem and Yelena Yesha)
________	TR-95-148	Unsupervised, Robust Estimation-based Clustering of Remotely Sensed Images (Nathan S. Netanyahu, James C. Tilton and J. Anthony Gualtieri)
________	TR-95-149	Knowledge Discovery from Structural Data (Diane J. Cook, Lawrence B. Holder and Surnjani Djoko)
________	TR-95-150	Online Data Compression in a Mass Storage File System (Odysseas I. Pentakalos and Yelena Yesha)
________	TR-95-151	A User's Guide to Pablo® I/O Instrumentation (Ruth A. Aydt)
________	TR-95-152	Input/Output Characteristics of Scalable Parallel Applications (Phyllis E. Crandall, Ruth A. Aydt, Andrew A. Chien, Daniel A. Reed)
________	TR-95-153	Towards a Parallel Registration of Multiple Resolution Remote Sensing Data (Jacqueline Le Moigne)
________	TR-95-154	PPFS: A High Performance Portable Parallel File System (James V. Huber Jr., Christopher L. Elford, Daniel A. Reed, Andrew A. Chien, David S. Blumenthal)
________	TR-95-155	An Approximate Performance Model of a Unitree Mass Storage System (Odysseas I. Pentakalos, Daniel A. Menasce, Milt Halem, Yelena Yesha)
________	TR-95-156	Communication Strategies for Out-of-Core Programs on Distributed Memory Machines (Rajesh Bordawekar and Alok Choudhary)
________	TR-95-157	Optimal Allocation for Partially Replicated Database Systems on Ring Networks (A. B. Stephens, Yelena Yesha and Keith Humenik)
________	TR-95-158	Minimizing Message Complexity of Partially Replicated Data on Hypercubes (Keith Humenik, Peter Matthews, A. B. Stephens and Yelena Yesha)
________	TR-95-159	Two Approaches for High Concurrency in Multicast-Based Object Replication (Theodore Johnson and Lionel Maugis)

________	TR-95-160	Designing Distributed Search Structures with Lazy Updates (Theodore Johnson and Padmashree Krishna)
________	TR-95-161	The Proceedings of The Petaflops Frontier Workshop-February 6, 1995 (Thomas Sterling and Michael J. MacDonald)
________	TR-95-162	Findings of the Second Pasadena Workshop on System Software and Tools for High Performance Computing Environments (Thomas Sterling, Paul Messina and Jim Pool)
________	TR-95-163	An Experimental Study of Input/Output Characteristics of NASA Earth and Space Sciences Applications (Michael R. Berry and Tarek El-Ghazawi)
________	TR-95-164	An Analytic Model of Hierarchical Mass Storage Systems with Network-Attached Storage Devices (Daniel A. Menasce, Odysseas I. Pentakalos and Yelena Yesha)
________	TR-95-165	Analytical Performance Modeling of Hierarchical Mass Storage Systems (Odysseas I. Pentakalos, Daniel Menasce, Milt Halem and Yelena Yesha)
________	TR-95-166	CESDIS Annual Report; Year 6
________	TR-95-167	CESDIS Annual Report; Year 7
________	TR-96-168	Evaluation of Segmented Ethernet Interconnect Topologies for the Beowulf Parallel Workstation (Chance Reschke, Thomas Sterling, Donald J. Becker, Daniel Ridge, Phillip Merkey, Odysseas Pentakalos and Michael R. Berry)
________	TR-96-169	The Performance of Earth and Space Science Applications on the Convex Exemplar Scalable Shared Memory Multiprocessor (Thomas Sterling, Phillip Merkey and Daniel Savarese)
________	TR-96-170	Parallel Input/Output Issues in Sparse Matrix Computations (Sorin G. Nastea, Tarek El-Ghazawi and Ophir Frieder)
________	TR-96-171	The Use of Wavelets for Remote Sensing Image Registration and Fusion (Jacqueline Le Moigne and Robert F. Cromp)
________	TR-96-172	I/O, Performance Analysis, and Performance Data Immersion (Daniel A. Reed, Tara Madhyastha, Ruth A. Aydt, Christopher L. Elford, Will H. Scullin, Evgenia Smirni)
________	TR-96-173	Optimal Allocation of Replicated Data in Tree Networks (A. B. Stephens, David M. Lazoff and Yelena Yesha)
________	TR-96-174	Analytical Performance Modeling of Hierarchical Mass Storage Systems (Odysseas I. Pentakalos, Daniel A. Menascé, Milt Halem, Yelena Yesha)
________	TR-96-175	Analytical Modeling of Distributed Hierarchical Mass Storage Systems with Network-Attached Storage Devices (Odysseas I. Pentakalos, Daniel A. Menascé, and Yelena Yesha)

________	TR-96-176	MIPI: Multi-level Instrumentation of Parallel Input/Output (Michael R. Berry and Tarek A. El-Ghazawi)
________	TR-96-177	Tuning the Performance of I/O-Intensive Parallel Applications (Anurag Acharya, Mustafa Uysal, Robert Bennett, Assaf Mendelson, Michael Beynon, Jeff Hollingsworth, Joel Saltz, Alan Sussman)
________	TR-96-178	Interactive Smooth-Motion Animation of High Resolution Ocean Circulation Calculations (Aaron Sawdey, Derek Lee, Thomas Ruwart, Paul Woodward, Matthew O'Keefe, Rainer Bleck)
________	TR-96-179	A comparison of data-parallel and message-passing versions of the Miami Isopycnic Coordinate Ocean Model (MICOM) (Rainer Bleck, Sumner Dean, Matthew O'Keefe, Aaron Sawdey)
________	TR-96-180	Instrumenting a Unix Kernel for Event Tracing (Steven R. Soltis, Matthew T. O'Keefe, Thomas M. Ruwart)
________	TR-96-181	An Object Oriented Performance Analyzer of Hierarchical Mass Storage Systems (Odysseas I. Pentakalos, Daniel A. Menascé, Yelena Yesha)
________	TR-96-182	An Automated Parallel Image Registration Technique of Multiple Source Remote Sensing Data (Jacqueline LeMoigne, William J. Campbell, Robert F. Cromp)
________	TR-96-183	Using High Performance Fortran for Earth and Space Science Applications (Terrence W. Pratt)
________	TR-96-184	Performance Evaluation of Piecewise Parabolic Method on Convex Exemplar SPP1000 (Udaya A. Ranawake)
________	TR-96-185	Accessing Sections of Out-of-Core Arrays Using an Extended Two-Phase Method (Alok Choudhary, Rajeev Thakur)
________	TR-96-186	A Visual Database System for Image Analysis on Parallel Computers and its Application to the EOS Amazon Project (Linda Shapiro, Steven Tanimoto, James Ahrens
________	TR-96-187	Image Categorization Using Texture Features (Aya Soffer)
________	TR-97-188	Pythia and Pythia/WK: Tools for the Performance Analysis of Mass Storage Systems (Odysseas Pentakalos, Daniel Menascé, Yelena Yesha)
________	TR-97-189	Analytical Modeling of Robotic Tape Libraries Using Stochastic Automata (Tugrul Dayar, Odysseas Pentakalos, A. B. Stephens)
________	TR-97-190	Electronic Commerce and Digital Libraries: Towards a Digital Agora (Nabil Adam, Yelena Yesha)
________	TR-97-192	Automated Clustering-Based Workload Characterization (Odysseas Pentakalos, Daniel Menascé, Yelena Yesha)
________	TR-97-193	An Empirical Evaluation of the Convex SPP-1000 Hierarchical Shared Memory System (Thomas Sterling, Daniel Savarese, Phillip Merkey, Kevin Olson)

________	TR-97-194	Globalizing Business, Education, Culture Through the Internet (Nabil Adam, Baruch Awerbuch, Jacob Slonim, Peter Wegner, Yelena Yesha)
________	TR-97-195	CESDIS Annual Report; Year 8
________	TR-97-196	Hubble Space Telescope Faint Object Camera Calculated Point-Spread Functions (Rick Lyon, Jan M. Hollis, John Dorband)
________	TR-97-197	Motion of the Ultraviolet R Aquarii Jet (Rick Lyon, Jan M. Hollis, John Dorband, W.A. Feibelman)
________	TR-97-198	A Maximum Entropy Method with a Priori Maximum Likelihood Constraints (Rick Lyon, Jan M. Hollis, John Dorband)
________	TR-97-199	Information Extraction Based Multiple-Category Document Classification for the Global Legal Information Network (Nabil Adam, Richard Holowczak)
________	TR-97-200	The Global Legal Information Network (GLIN) (Nabil Adam, Burt Edelson, Tarek El-Ghazawi, Milt Halem, Kostas Kalpakis, Nick Kozura, Rubens Medina, Yelena Yesha)
________	TR-97-201	Modeling and Analysis of Workflows Using Petri Nets (Nabil Adam, Vijayalakshmi Atluri, Wei-Kuang Huang)
________	TR-97-202	Pythia: A Performance Analyzer of Hierarchical Mass Storage Systems (Odysseas Pentakalos, Daniel Menascé, Yelena Yesha)
________	TR-97-203	Wavelet-Based Image Registration on Parallel Computers (Tarek El-Ghazawi, Prachya Chalermwat, Jacqueline Le Moigne)
________	TR-97-204	An Architecture-Independent Workload Characterization Model for Parallel Computer Architectures (Abdullah Ibrahim Meajil) Dissertation
________	TR-97-205	Project Management - Industrial Prospective: Focusing on Industrial Software Engineering; Best Practice for Developing High (Jacob Slonim)
________	TR-97-206	Proceedings of the Image Registration Workshop (Jacqueline Le Moigne)
________	TR-97-207	Satellite Imaging and Sensing (Jacqueline LeMoigne, Robert Cromp)
________	TR-97-208	Accessing Sections of Out-of-Core Arrays Using an Extended Two-Phase Method (Alok Choudhary, Rajev Thakur)
________	TR-97-209	Efficient Compilation of Out-of-Core Data Parallel Programs (Alok Choudhary, Rajesh Bordawekar, Rajeev Thakur)
________	TR-97-210	The Design of VIP-FS: A Virtual, Parallel File System for High Performance Parallel and Distributed Computing (Alok Choudhary, Juan Miguel del Rosario, Michael Harry)
________	TR-97-211	Runtime Support for Parallel I/O in PASSION (Alok Choudhary, Rajeev Thakur, Rajesh Bordawekar, Ravi Ponnusamy)

________	TR-97-212	Architecture-Independent Locality-Improving Transformations of Computational Graphs (Sanjay Ranka, Chao-Wei, Manoj Gunwani)
________	TR-97-213	SPRINT: Scalable Partitioning, Refinement, and Incremental Partitioning Techniques (Sanjay Ranka, Chao-Wei Ou)
________	TR-98-214	An Assessment of Beowulf-class Computing for NASA Requirements: Initial Findings from the First NASA Workshop on Beowulf-class Clustered Computing (Don Becker, Thomas Sterling, Mike Warren, Tom Cwik, John Salmon, Bill Nitzberg)
________	TR-98-215	CESDIS Annual Report; Year 9
________	TR-98-216	Analysis of the Eigenstructure of the Chew, Goldberger and Low System of Equations (Dinshaw Balsara, Daniel Spicer)
________	TR-98-217	Maintaining Pressure Positivity in Magnetohydrodynamic Simulations (Dinshaw Balsara, Daniel Spicer)
________	TR-98-218	A Staggered Mesh Algorithm Using High Order Godunov Fluxes to Ensure Solenoidal Magnetic Fields in Magnetohydrodynamic Simulations (Dinshaw Balsara, Daniel Spicer)
________	TR-98-219	Achieving Ten Gflops on PC Clusters: A Case Study (Udaya Ranawake, John Dorband, Bruce Fryxell, Daniel Ridge, Erik Hendriks, Donald Becker, Phillip Merkey)
________	TR-98-220	Seamlessly Selecting the Best Copy from Internet-Wide Replicated Web Servers (Yair Amir, Alec Peterson, David Shaw)
________	TR-98-221	An Evaluation of Automatic Image Registration Methods (Jacqueline LeMoigne, Wei Xia, James Tilton, Prachya Chalermwat, Tarek El-Ghazawi, Nathan Netanyahu, David Mount, William Campbell)
________	TR-98-222	Parallel Knowledge Discovery from Large Complex Databases (Diane Cook, Lawrence Holder)

NAME____

ADDRESS_____________________________________

____

____

PHONE_______________________________________

E-MAIL______________________________________

AREAS OF RESEARCH INTEREST__________________

Appendix F

CESDIS Personnel and Associates

CESDIS ADMINISTRATIVE OFFICE

301-286-4403 fax: 301-286-1777

Individual extensions will roll over to another number or phonemail, if the party called does not answer. Please allow sufficient rings for this to happen.

U. S. Mail Address

CESDIS
Code 930.5
NASA Goddard Space Flight Center
Greenbelt, MD 20771

Federal Express/UPS Address

CESDIS
Building 28, Room W223
NASA Goddard Space Flight Center
Greenbelt, MD 20771

DIRECTOR

Yelena Yesha	UMBC: 410-455-3542	yeyesha@cs.umbc.edu

ACTING ASSOCIATE DIRECTOR

Susan Hoban	301-286-7980	shoban@pop900.gsfc.nasa.gov

SENIOR AND STAFF SCIENTISTS

Donald Becker	301-286-0882	becker@cesdis.edu
Jacqueline Le Moigne	301-286-8723	lemoigne@nibbles.gsfc.nasa.gov
Les Meredith	301-286-8830	les@usra.edu
Phillip Merkey	301-286-3805	merk@cesdis.edu
Terry Pratt	301-286-0880	pratt@cesdis.edu

TECHNICAL PERSONNEL

Chang-Hong Chien	301-286-0881	cchien@cesdis.edu
Erik Hendriks	301-286-0065	hendriks@cesdis.edu

ADMINISTRATIVE PERSONNEL

Nancy Campbell	301-286-4099	campbell@cesdis.usra.edu
L'Tanya Clark	301-286-8951	clark@cesdis.usra.edu
Georgia Flanagan	301-286-2080	georgia@cesdis.usra.edu
Michele Meyett	301-286-3062	shelly@cesdis.usra.edu
Dawn Segura	301-286-0913	dawn@cesdis.usra.edu
Yolanda Smith	301-286-8755	smith@cesdis.usra.edu

UNIVERSITY/INDUSTRY PROJECT PERSONNEL

Brandeis University
Computer Science Department
Waltham, MA 02254-9110

James Storer	617-736-2714	storer@cs.brandeis.edu

Bucknell University
Department of Electrical Engineering
Lewisburg, PA 17837

Maurice Aburdene	717-524-1234	aburdene@bucknell.edu

Fran Stetina and Associates
Bowie, MD 20715

Fran Stetina	301-286-0769	fran@suzieq.gsfc.nasa.gov

George Mason University
Department of Information and Software Engineering (ISE)
Fairfax, VA 22030

Alexander Brodsky	703-993-1529	brodsky@gmu.edu

George Mason University
Institute for Computational Science and Informatics
Fairfax, Va 22030

Tarek El-Ghazawi	CESDIS: 301-286-8178 GMU: 703-993-3610	tarek@science.gmu.edu

George Mason University
Department of Computer Science
Fairfax, VA 22030-4444

Daniel Menascé	703-993-1537	menasce@cs.gmu.edu

George Washington University
Institute for Applied Space Research
Washington, DC 10052

Burt Edelson	202-994-5509	edelson@seas.gwu.edu

Georgia Institute of Technology
Broadband and Wireless Networking Laboratory
Atlanta, GA 30332

Ian Akyildiz	404-894-5141	ian@ee.gatech.edu

The Johns Hopkins University
Department of Computer Science
Baltimore, MD 21218

S. Rao Kosaraju	410-516-8134	kosaraju@cs.jhu.edu
Gregory Solyar	410-486-4552	solyar@doomsday.ece.jhu.edu

Northwestern University
Department of Electrical and Computer Engineering
Evanston, IL 60208

Bin Chen	847-491-7141	bchen@ece.nwu.edu

Rutgers University
Center for Information Management, Integration, and Connectivity
180 University Avenue
Newark, NJ 07102

Nabil Adam	973-353-5239	adam@adam.rutgers.edu

Scripps Institution of Oceanography
University of California, San Diego
La Jolla, CA 92093-0224

Richard Somerville	619-534-4644	rsomerville@ucsd.edu

University of California
Computer Engineering Department
Room 225, Applied Sciences
Santa Cruz, CA 95064

Glen Langdon	408-459-2212	langdon@cse.ucsc.edu

University of Illinois at Chicago
Department of Electrical Engineering and Computer Science
Chicago, IL 60607

Ouri Wolfson	312-996-6770	wolfson@ouri.eecs.uic.edu

University of Maryland Baltimore County

Department of Computer Science and Electrical Engineering
5401 Wilkens Avenue
Baltimore, MD 21228-5398

Kostas Kalpakis	410-455-3143	kalpakis@cs.umbc.edu
Richard Lyon	301-286-4302	lyon@jansky.gsfc.nasa.gov
Tim Murphy	301-286-9805	murphy@albert.gsfc.nasa.gov
Udaya Ranawake	301-286-3046	udaya@neumann.gsfc.nas.gov
Miodrag (Misha) Rancic	301-286-2439	mrancic@ciga.gsfc.nasa.gov
Joel Sachs	410-455-6338	jsachs@cs.umbc.edu
Sushel Unninayar	301-286-2757	sushel@cesdis.edu

University of Maryland College Park

Department of Computer Science
College Park, MD 20742

Joel Saltz	301-405-2684	saltz@cs.umd.edu

University of North Carolina

Department of Computer Science
Sitterson Hall
Chapel Hill, NC 27599-3175

David Stotts	919-962-1833	stotts@cs.unc.edu

University of Rochester

Department of Physics and Astronomy
Bausch and Lomb Building
Rochester, NY 14627-0171

Adam Frank	716-275-1717	afrank@alethea.pas.rochester.edu

University of Washington

Department of Astronomy
FM-20
Seattle, WA 98195

George Lake	206-543-7106	lake@astro.washington.edu

University of Waterloo

Department of Computer Science
Waterloo, Ontario N2L 3G1, Canada

Kenneth Salem	519-888-4567, ext 3485	kmsalem@zonker.uwaterloo.ca

ALPHABETICAL DIRECTORY OF CESDIS PERSONNEL

Adam, Nabil	973-353-5239	adam@adam.rutgers.edu
Akyildiz, Ian	404-894-5141	ian@ee.gatech.edu
Becker, Don	301-286-0882	becker@cesdis.usra.edu
Campbell, Nancy	301-286-4099	campbell@cesdis.usra.edu
Chien, Chang-Hong	301-286-0881	cchien@cesdis.usra.edu
Clark, L'Tanya	301-286-8951	clark@cesdis.usra.edu
de La Beaujardière, Jeff	301-286-1569	delabeau@iniki.gsfc.nasa.gov
Edelson, Burt	202-994-5509	edelson@seas.gwu.edu
El-Ghazawi, Tarek	703-993-3610	tarek@science.gmu.edu
Flanagan, Georgia	301-286-2080	georgia@cesdis.usra.edu
Frank, Adam	716-275-1717	afrank@alethea.pas.rochester.edu
Gurski, Katharine	301-286-9096	kgurski@cesdis.usra.edu
Hendriks, Erik	301-286-0065	hendriks@cesdis.edu
Hoban, Susan	301-286-7980	shoban@pop900.gsfc.nasa.gov
Kosaraju, S. Rao	410-516-8134	kosaraju@cs.jhu.edu
Lake, George	206-543-7106	lake@astro.washington.edu
Le Moigne, Jacqueline	301-286-8723	lemoigne@nibbles.gsfc.nasa.gov
Lyon, Richard	301-286-4302	lyon@jansky.gsfc.nasa.gov
Menascé, Daniel	703-993-1537	menasce@cs.gmu.edu
Meredith, Les	301-286-8830	les@usra.edu
Merkey, Phillip	301-286-3805	merk@cesdis.usra.edu
Meyett, Michele	301-286-4403	shelly@cesdis.usra.edu
Murphy, Tim	301-286-9805	murphy@albert.gsfc.nasa.gov
Norris, Peter	301-286-2081	pnorris@cesdis.gsfc.nasa.gov
Pratt, Terry	301-286-0880	pratt@cesdis.usra.edu
Ranawake, Udaya	301-286-3046	udaya@neumann.gsfc.nasa.gov
Rancic, Miodrag (Misha)	301-286-2439	mrancic@ciga.gsfc.nasa.gov
Sachs, Joel	410-455-6338	jsachs@cs.umbc.edu
Segura, Dawn	301-286-0913	dawn@cesdis.usra.edu
Smith, Yolanda	301-286-8755	smith@cesdis.usra.edu
Solyar, Greg	410-486-4552	solyar@doomsday.ece.jhu.edu
Somerville, Richard	619-534-4644	rsomerville@ucsd.edu
Stetina, Fran	301-286-0769	fran@suzieq.gsfc.nasa.gov
Unninayar, Sushel	301-286-2757	sushel@cesdis.usra.edu
Yesha, Yelena	UMBC: 410-455-3542	yeyesha@cs.umbc.edu