Mitsubishi Electric Research Laboratories (MERL)

Annual Report

July 2000 through June 2001

TR2001-00

Welcome to MERL – Mitsubishi Electric Research Laboratories, the North American corporate R&D arm of Mitsubishi Electric Corporation (MELCO). In this report you will find descriptions of MERL as a whole, our three laboratories (MERL Cambridge Research, MERL Cambridge Systems, and MERL Murray Hill), and most importantly our current projects.

201 Broadway, Cambridge, Massachusetts 02139
617.621.7500

Table of Contents

Mitsubishi Electric Research Laboratories

Mitsubishi Electric Research Laboratories (MERL) is the North American arm of the central research and development organization of Mitsubishi Electric Company (MELCO). MERL conducts application-motivated basic research and advanced development in computer and communications technology.

MERL's mission—our assignment from MELCO—is twofold.

- To generate highly significant intellectual property (papers, patents, and prototypes) in areas of importance to MELCO.
- To locate organizations within MELCO that can benefit from this technology and deliver it to them in forms they can use effectively (specifications, software components, and systems).

MERL's vision—our goal for ourselves—is also twofold.

- To be one of the world's premiere research laboratories, significantly advancing the frontiers of technology and making lasting impacts on the world.
- Within our areas of expertise, to be the prime source of new technology for MELCO.

MERL focuses on three key technology sectors: (1) human/computer interaction (HCI) featuring computer vision, computer graphics & speech interfaces, (2) Internet computing & applications, and (3) digital communications featuring video processing & networking. The labs within MERL focus on specific segments of these technologies, while working collaboratively to achieve groundbreaking results. Our output ranges from papers and patents, through proof-of-concept hardware and software prototypes, to industry-first products.

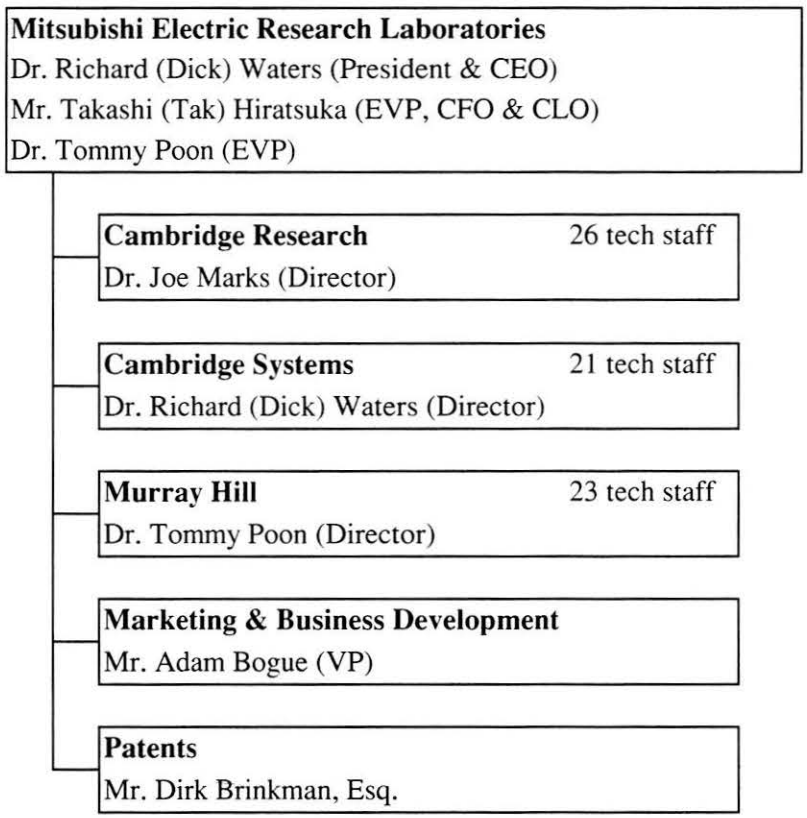
MERL is small enough to be agile and flexible in the dynamic marketplace of ideas. In addition, we gain leverage from the size, recognition and diversity of our strong global parent. We turn our technical achievements into business successes by partnering with MELCO's business units and with other labs in MELCO's global R&D network.

We are strongly involved in the R&D community and standards activities, maintaining long-standing cooperative relationships with research universities such as Brown, CMU, Georgia Tech, MIT, the University of Michigan, and Princeton. We encourage our staff to be involved in their professional communities via conferences, papers, and continuing professional development.

As shown in the chart on the next page, MERL currently consists of three laboratories, one in Murray Hill New Jersey and two in Cambridge Massachusetts. (See the lab description sections that follow.) MERL's headquarters (also in Cambridge) includes a small marketing and business development department to help assess the potential market impact of our work and an in-house patent department to speed the filing of patents.

A fourth MERL laboratory was located in Concord MA. Over the past two years the focus of this lab shifted from advanced development to development of a volume rendering chip. In February, 2001, the Concord lab left MERL becoming part of the business selling the chip.

This annual report is a snapshot in time of MERL's web site. For additional and updated information please visit our web site at "http://www.merl.com".



Richard (Dick) Waters
Ph.D., Massachusetts Institute of Technology, 1978
President & Chief Executive Officer; Interim Director Cambridge Systems

Dick Waters received his Ph.D. in artificial intelligence (AI). For the next 13 years he worked at the MIT AI Lab as a Research Scientist and co-principal investigator of the Programmer's Apprentice project. Dick was a founding member of MERL Cambridge Research in 1991. As a research scientist, his work centered on multi-user interactive environments for work, learning and play. In January 1998, Dick became director of MERL Cambridge Research. In December 1999, he became CEO of MERL as a whole. (At the time of the printing of this report, Dick is acting as the Interim Director of MERL Cambridge Systems during the search for a permanent Director.)

Takashi (Tak) Hiratsuka
B.S., Univ. of Tokyo, 1969
Executive Vice President, Chief Financial Officer & Chief Liaison Officer

After many years in MELCO's computer business unit, Tak Hiratsuka came to the US in 1988, as the liaison between MELCO and MELCO's subsidiary Horizon Research Inc. After 6 years there, Tak took on other duties for MELCO including a stint as Assistant Deputy Director of Corporate R&D in Tokyo. In December 1999, Tak returned to the US as MELCO's Liaison to MERL. Tak plays an essential role in building partnerships between MERL and MELCO.

Tommy Poon
Ph.D., Columbia University, 1980
Executive Vice President; Director Murray Hill

After receiving his Ph.D., Tommy Poon worked briefly at RCA Laboratory before moving to AT&T Bell Laboratories. During 13 years with Bell Labs, Tommy managed numerous R&D projects in telecommunications and signal processing. He joined MERL Murray Hill as Director in 1995. His primary research interests include digital and wireless communications, digital video and digital networks.

Joe Marks
Ph.D., Harvard University, 1991
Vice President; Director Cambridge Research

Prior to Joining MERL in 1994, Joe Marks worked at Digital Equipment Corporation's Cambridge Research Laboratory. As a researcher at MERL, Joe's primary focus was on computer graphics, user interfaces, and heuristic optimization. Joe also has a strong interest in teaching. He was an adjunct lecturer in the Division of Engineering and Applied Sciences at Harvard University from 1991 to 1996. In 1999 Joe became associate director of MERL Cambridge Research. In late 2000, he replaced Dick Waters as director.

Adam Bogue
B.S., MIT, 1986; MBA, MIT Sloan School, 1990
Vice President of Marketing and Business Development

For 3 years at GenRad Inc, Adam Bogue was responsible for managing a new line of automatic test equipment. Prior to joining MERL, Adam spent the 7 years at Active Control eXperts Inc. beginning as Director of Sales and Marketing and ending as Vice President, Core and New Business Unit. Adam began work for MERL in June of 2000.

Dirk Brinkman, Esq.
M.Sc., Univ. of Toronto, 1970; J.D., Suffolk Univ. Law School, 1990
Patent Counsel

Dirk Brinkman's undergraduate and Masters work was in Medical Physics. Prior to joining MERL in 1998, he pent most of his career at Digital Equipment Corporation, first as an engineer and product manger in the Medical Systems Group, and then as a Patent Attorney for Digital's Research Laboratories in Cambridge MA and Palo Alto CA.

Mitsubishi Electric

Number 98 on Fortune magazine's list of the world's 500 largest corporations, Mitsubishi Electric Corporation (MELCO) has approximately $33 billion in annual sales and 117,000 employees in 34 countries. Like most Japanese companies, the lingering malaise of the Japanese economy has affected MELCO leading to losses in 1997 and 1998. However, MELCO returned to profitability in 1999 and achieved the largest profits in its history in 2000.

MELCO is organized into nine business units (listed below in order of revenues). Because information technology is important to each of these business units, MERL works with them all. (The rightmost column below shows the abbreviated Japanese business unit names commonly used by MELCO insiders.)

Mitsubishi Electric Diversified Electrical and Electronics Manufacturer		MELCO
Living Environment & Digital Media Equip Air Conditioners, Refrigerators, TVs, DVDs, LCD Projectors	(Shizuoka, Kyoto)	Lihon
Semiconductors Memory, Microcontrollers, Optical Semiconductors, LCD Panels	(Kita Itami)	Hanpon
Building Systems Elevators, Escalators, Building Monitoring	(Inazawa)	Biruhon
Social Infrastructure Systems Power Equipment, Plant Control, Transportation	(Kobe, Itami)	Shakaihon
Electronic Systems Satellites, Radar, Military Systems	(Kamakura, Itami)	Denshihon
Communication Systems Wired Communications, Broadcast Communications, Cell Phones	(Kamakura, Itami)	Tsuhon
Automotive Equipment Alternators, Engine Controllers, Car Stereos, Car Navigation	(Himeji, Sanda)	Shahon
Factory Automation Programmable Logic Controllers, Industrial Machine Tools	(Nagoya)	FAhon
Information Systems and Services Turnkey Information Systems, Computer Hardware, Networking services	(Tokyo, Kamakura)	IShon

It is worthy of note that there are over 30 major independent companies that use the word "Mitsubishi" in their names. These companies include the Mitsubishi trading company,

Mitsubishi Chemicals, Mitsubishi Motors, Mitsubishi Heavy Industries, Mitsubishi Materials, and Bank of Tokyo-Mitsubishi (all six of which are also on the Fortune Global 500 list). They have shared roots in 19th century Japan; however, these companies have been separate for many years and MELCO has been separate from all of them since its founding in 1921.

Many of MELCO's business units have North American subsidiaries. These subsidiaries, particularly those that do design as well as manufacturing and sales, are natural partners of MERL. The subsidiaries that have (or are soon expected to have) sales of over one hundred million dollars per year are listed below in order of sales volume.

Mitsubishi Electric United States, Inc. (MEUS)
Sales: Several BUs (Los Angeles, Sunnyvale & other cities)
Semiconductors, Air Conditioning, Elevators

Mitsubishi Digital Electronics America, Inc. (MDEA)
Design, Manufacturing & Sales: Lihon (Los Angeles, Mexicali MX)
High Definition Projection Televisions, DVDs, VCRs

Mitsubishi Electric Automotive America, Inc. (MEAA)
Manufacturing & Sales: Shahon (Detroit, Mason OH)
Auto Parts

Mitsubishi Wireless Communications, Inc. (MWCI)
Design, manufacturing & Sales: Tsuhon (Atlanta, San Diego)
Cell Phones

Mitsubishi Electric Sales Canada, Inc. (MESCA)
Sales: Several BUs (Toronto & other cities)
Semiconductors, Air Conditioning, Projectors

Mitsubishi Electric Display Devices America, Inc. (MDDA)
Manufacturing & Sales: Lihon (Calexico CA, Mexicali MX)
Auto Parts

Mitsubishi Electric Automation, Inc. (MEAU)
Sales & Installation: FAhon (Chicago)
Factory Automation Equipment

Mitsubishi Electric Power Products, Inc. (MEPPI)
Design, Manufacturing & Sales: Shakaihon (Pittsburgh)
Power Transmission Products

Paceon Corporation (PACEON)
Design, Manufacturing & Sales: Tsuhon (Atlanta)
ATM Passive Optical Networks

Mitsubishi Electric Corporate R&D

MERL is one of eight laboratories in MELCO's Global Corporate R&D network. The chart below summarizes the primary activities of these labs. MERL pursues collaborations with all these labs. (The rightmost column shows the Japanese nicknames commonly used by insiders.)

Corporate R&D — Hatsuhon
Headquarters: Dr. Hitoshi Ogata (director), Dr. Hiroshi Koezuka (GM), 25 people (Tokyo)
Managing MELCO's R&D

- **Advanced Technology R&D Center (ATC)** — Sentansoken
  Research & Advanced Development: Dr. Tadatoshi Yamada (GM), 650 people (Itami)
  Materials, Semiconductor Devices, Electrical & Mechanical Engineering
- **Information Technology R&D Center (ITC)** — Johosoken
  Advanced Development: Mr. Shuichi Ono (GM), 800 People (Ofuna)
  Information Systems, Communications, Opto-Electronics
- **Industrial Electronics & Systems Laboratory (IESL)** — Sanken
  Advanced Development: Mr. Mitchitaka Ohshima (GM), 260 people (Itami)
  Power Electronics, Plant Control, Factory Automation, Building Systems
- **Industrial Design Center (IDC)** — IDken
  Advanced Development: Mr. Seiji Wada (GM), 70 people (Ofuna)
  Industrial Design, Usability Studies
- **Imaging Systems Laboratory (ISL)** — Eijoken
  Advanced Development: Mr. Takehiko Tsuiki (GM), 120 people (Kyoto)
  CRTs, Projection TVs, Digital receivers, DVD players
- **Information & Communications Systems Development Center (ICSD)** — Jotsuken
  Advanced Development: Mr. Kenji Kuroda (GM), 120 people (Kamakura)
  Information systems, Networked systems, Internet services
- **Mitsubishi Electric Research Laboratories (MERL)** — MERL
  Research & Advanced Development: Dr. Richard Waters (CEO), 90 people (MA & NJ)
  Computer Vision, HCI, Internet Software, Digital Audio & Video Communications
- **Mitsubishi Electric Information Technology Centre Europe (ITE)** — ITE
  Advanced Development: Mr. Ryugo Nishii (CEO), 60 people (France & England)
  Wireless Communications, Digital Audio & Video Broadcasting

MERL Cambridge Research

MERL Cambridge Research pursues basic research in applied computing. Our efforts are directed towards applications of practical significance, but our time horizon is long (five or more years) and our appetite for technical challenge and risk is high. Located in Cambridge, Massachusetts, hometown of Harvard University and the Massachusetts Institute of Technology, the laboratory currently has a technical staff of 26. The permanent staff is enriched by an active program of student internships hosting approximately 40 students per year for an average of three to four months each.

The best ideas are born and mature most quickly when critically examined and refined by many minds. Toward this end, MERL researchers are encouraged to work together with each other and with researchers at other institutions. In particular we forge collaborations that carry our technologies into products by connecting to our sister organizations in MERL and MELCO. Further, we participate actively in external research communities, exposing our work to critical peer review, and publishing our results as quickly as possible.

To allow rapid response to opportunities, MERL Cambridge Research is organized as a flexible community of researchers without internal divisions. Our goal is to support a continually changing mix of individual explorations and group projects, where promising individual efforts can easily grow into projects and projects can easily disband at the end of their natural lives. Each researcher is typically involved in both individual explorations and related group projects.

Over the last few years, perceptual user interaction (PUI) has emerged as a central theme of the lab. This theme is providing a unifying focus for our research in computer graphics, computer vision, audio and image processing, natural-language processing, speech processing, machine learning, networked and wireless communication, and human-computer interaction. Concretely, this has generated projects in such important business areas as video surveillance, telephony, and consumer electronics. The PUI theme is resonating well with MELCO labs and business units, producing better connections for our research than ever before.

Yet while we are strengthening our connections with MELCO's business interests, we are also improving our record of scientific and technological achievement, as measured by our recent publications and standards activities. Within its areas of expertise, MERL Cambridge Research is widely regarded as one of the premier industrial research laboratories in the world.

Successful basic research in an industrial setting requires many things: a world-class team of researchers; a long-term commitment to open-ended, risky exploration; research themes that resonate well with the parent company; and effective tech-transfer mechanisms whereby research ideas can be converted into new products. These elements are now all in place at MERL Cambridge Research, so we have good reason to look to the coming years with confidence.

MERL Cambridge Systems

MERL Cambridge Systems does technology integration and advanced development in four areas: Agent technology, Network services, Surveillance and human-computer interaction, and Speech interfaces.

Agent technology focuses on the development of software tools that can act autonomously and proactively on behalf of some entity, either human or software. They are especially useful to off-load, from humans to computers, drudgery that is tedious, exhausting, and time-consuming – conditions ripe for human error. Our pure Java mobile agent system, Concordia, has been delivered to MELCO in both Java 1 and Java 2 versions. MELCO renamed it DiaConcord, and fielded it as a unique and flexible programming paradigm for distributed systems, for customers in a power system management application. We are evaluating both extensions to Concordia, and development of other agent system technologies, in collaboration with other MELCO labs.

Network services focuses on creating leading-edge software to deliver LAN-based and internet-based services. One of these is Net Rep, a joint venture with Veritas Software to build large scale, fault-resilient software systems to perform database replication/backup over a network; in 2001, MERL's licensing revenues from this product are expected to approach $1M. Another project, the Collaborative Web Browser, promotes joint use of web resources by letting multiple Internet Explorer users share the same web page, and focus on specific information in it.

In Surveillance and human-computer interaction, CSL is serving as the technology integrator for the MERL-wide effort (CSL, CRL, MHL) to create both applications that will aid surveillance, and a modular architecture that allows of tight system integration, while facilitating the hardware and software updates necessary to keep the system current and robust. We hope to work with many MELCO business units to support MELCO's substantial, distributed effort in the growing surveillance market, by creating vision-based software modules that provide unique capabilities as well as software differentiation. Other human-computer interaction modes were explored with the BufferPhone project, a clever simplification of the phone use scenario, and with DiamondTouch, a rugged touch-sensitive collaborative work/play surface.

Speech interfaces aims both at first-generation products in which speech is added to an already existing interface to improve the overall user capability (e.g. DiamondTalk), as well as second-generation products that are only conceivable and possible in the presence of a speech interface (e.g. One-Button Phone). This is a major shift of emphasis, in which CSL seized an opportunity and hired a number of outstanding individual contributors, resulting in a world-class team for speech applications research and development. This leverages the facts that MERL has a long history in user interfaces, that speech is emerging as the next major interface, and that the USA and Boston are centers of excellence in speech technology . We hope to work many MELCO business units to develop interfaces and product ideas, particularly in the communication, consumer goods, and automotive areas.

MERL Murray Hill

MERL Murray Hill is structured to address three major areas of research and advanced technology development: Digital Video, Digital Communications and Digital Networks.

The Digital Video Technology (DVT) group is developing techniques for digital video compression, communication and multimedia applications. Their effort is directed towards the development of an improved HDTV receiver, down-conversion decoding algorithms, video-coding algorithms, video transcoding algorithm, video segmentation algorithm, and video description, indexing and retrieval applications. DVT is an active participant and contributor to the MPEG standards including MPEG-4, MPEG-7 and MPEG-21.

The Digital Communications Group is focusing on the research and development of leading edge technologies for high speed digital communications systems with high quality of services. Our main research areas include next generation wireless communications systems (e.g. WCDMA and HSDPA), wireless local area networks, high speed Ethernet and optical networks, and advanced modem technologies. One of the primary goals of this group is to create new concepts and technologies, develop new algorithms, new designs that will enhance the QoS of wireless communications networks and high speed Ethernet that will enrich MELCO's IPR portfolio. In addition, the team is actively participating in various industry standards development, and collaborating with MELCO's business units to develop competitive communications products.

The Advanced Digital Networks (ADN) group is focusing on the design and integration of key technologies to link broadband and narrowband external networks to local (and/or personal) networks based on wireless and wireline physical layers. By taking advantage of key open protocols such as Internet Protocol, MPEG, IEEE1394, and others; a wide variety of end-to-end usage models supporting QoS, plug-and-play, guaranteed delivery, mobility, and security are being investigated. ADN is actively participating in and contributing to various industry standards related to Home Networking, Digital Broadcasting, and Consumer Electronics.

MERL Murray Hill has filed more than seventy patents arising from the above activities. Its expertise includes: 1) System knowledge of digital video-coding algorithms, digital communications systems and algorithms, wireless and wireline communications, digital broadcasting and data-networking software, and multiple-standard modem technologies.; 2) Standards in Industry MPEG-2, MPEG-4, MPEG-7, XDSL, TIA-3GPP2 & 3GPP; consumer electronics: ATSC, EIA CEA, ATTC & IEEE; home networking: 1394TA & UPnP.; 3) Fast prototyping of real-time hardware and firmware, digital signal processing, data-acquisition techniques and use of field-programmable gate array (FPGA) devices.; and 4) Fast prototyping of real-time hardware and firmware, digital signal processing, data-acquisition techniques and use of field-programmable gate array (FPGA) devices.

MERL Murray Hill has developed close working relationships with four MELCO R&D labs: Mitsubishi Digital Equipment America (digital projection TV with 1394 interface), the Information Technology R&D Center (advanced digital broadcasting), the System LSI Development Center (high speed circuits), and the Communication Systems R&D Center (wireless system). MERL Murray Hill and MELCO's Semiconductor Business Unit teamed with Lucent Technology on the design of the world's first digital HDTV receiver chip set. The laboratory is located in Murray Hill NJ, with easy access to New York City and major airports.

Technical Staff

David Anderson
M.Sc., Carnegie Mellon University, 1995
Senior Research Scientist, MERL Cambridge Research

David Anderson is most interested in ideas that combine technical innovation, artistic vision, and business opportunity. Before joining MERL, he worked on the Mach and Andrew systems projects at CMU, and taught at the Pennsylvania Governor's School for the Sciences. At MERL, he has been part of creating the Diamond Park shared virtual world, the SPLINE platform for distributed multimedia, and Open Community. More recently, he has been developing applications that use tangible and perceptual interfaces. David Anderson also serves on the board of directors of the Web3D Consortium.

Paul Beardsley
Ph.D., Oxford University, 1992
Research Scientist, MERL Cambridge Research

Paul Beardsley is a computer vision researcher. He obtained a DPhil from Oxford for work on motion recovery and camera calibration using projective geometry. This was followed by a post-doc on navigation through unknown environments, sensing the environment with uncalibrated cameras undergoing unknown motion. His recent work has been on face tracking, and he is currently doing advanced development work on a system for extracting 3D face models using stereo cameras.

Ghulam Bhatti
Ph.D., Boston University, 1998
Member Technical Staff, MERL Murray Hill

Ghulam Bhatti received his Ph.D. in Computer Science, specializing in Distributed and Parallel Discrete Event Simulation. He joined the MHL in November 2000. Previously, Ghulam has worked as a Sr. Software Engineer at Evare LLC, Inc, developing software for a network switch, named as Winternet, and implementing an RSA cryptographic scheme. He also worked at Excel Tech. Ltd. (XLTEK) as an Embedded Software Engineer, developing embedded software for a portable EEG device. Currently, Ghulam is working on Home Networking and Digital TV. His interests include algorithms, embedded software development, and networking.

Emmanuelle Bourrat
M.S., Université de Rennes1, 1999
Member Technical Staff, MERL Cambridge Systems

Emmanuelle Bourrat graduated in 1999 from IRISA (Institut de Recherche en Informatique et Systèmes Aléatoires), Université de Rennes1, France. She earned a Master's degree in computer science and digital images. She was an intern at MERL from February 2000 to August 2000, where she worked on a tool to visualize and segment scans of the head. Then she joined CSL as a full-time employee in August 2000. Her main interest is in computer vision and medical imaging. She is currently working on the CHO (Computer Human Observation) project.

Matthew Brand
Ph.D., Northwestern University, 1994
Research Scientist, MERL Cambridge Research

Matthew Brand's work focuses on learning how the world behaves from sensory data. One goal is to make machines that steadily become better at interpreting, assisting, and even mimicking human activity. Recent results include an entropy optimization framework for learning and methods for synthesizing high-quality virtual data from probability models. These techniques are being demonstrated in "digital puppets" -- systems that learn to synthesize realistic, expressive human behavior. In 1999 Brand was named one of 100 top innovators of his generation by Technology Review.

Michael Casey
Ph.D., Massachusetts Institute of Technology, 1998
Research Scientist, MERL Cambridge Research

Michael Casey's research focus is in developing novel signal processing and pattern recognition techniques for analysis and synthesis of audio. For his thesis Michael developed a new audio analysis technique based on independent components analysis that makes possible perceptually salient feature decomposition of textured and noise-based sounds and sound scenes. As an intern at MERL Cambridge Research before completing his degree, he also participated in the development of audio software for SPLINE and produced the sound design for Diamond Park, a large-scale multi-participant virtual environment.

Johnas Cukier
M.Sc., from Polytechnic Institute of New York, 1985
Principal Member Technical Staff, MERL Murray Hill

Johnas Cukier received his B.Sc. degree in Physics and Computer Science in 1983 from New York University and his M.Sc. degree in Electrical Engineering in 1985 from Polytechnic Institute of New York. He joined MERL Murray Hill in 1996, working on digital systems for CATV, RF microwave transmitters/receivers, and front-end advanced TV receivers. His current interests are in advanced Digital Networking and Digital Signal Processing.

Andrea DeDuck
B.S.E.E., Rensselaer Polytechnic Institute, 1984
Principal Technical Staff, MERL Cambridge Systems

Andrea DeDuck has worked on numerous projects in her years at MERL Cambridge Systems as a CPU micro-code developer and systems verification team leader. Most recently she was project leader on the Netrep project, which produced MERL Cambridge Systems' first commercially licensed technology. In her spare time, Andrea likes to visit exotic places above and below sea level.

Thierry Derand
M.Sc., Ecole Superieure d'Ingenieurs en Electrotechnique et Electronique, 2000
Member Technical Staff, MERL Murray Hill

Thierry Dérand received a French Engineering diploma equivalent to M.Sc. degree in Electrical Engineering in 2000. He graduated from a French college, E.S.I.E.E. (Ecole Supérieure d'Ingénieurs en Electrotechnique et Electronique), Paris, France. He joined MERL Murray Hill in 2000, and work on 3G CDMA systems focusing on Diversity Techniques and Interference Cancellation. His main research interests are in general wireless communications.

Paul Dietz
Ph.D., Carnegie Mellon University, 1995
Principal Technical Staff, MERL Cambridge Systems

Paul Dietz is an over-educated, academic refugee who seeks happiness through creating clever devices and systems. Most recently, Paul headed up the electrical engineering efforts at Walt Disney Imagineering's Cambridge R&D lab where he worked on a wide variety of projects including theme park attractions, systems for the ABC television network and consumer products. Since joining MERL Cambridge Systems last year, Paul has been leading efforts developing new user interface technologies.

Ajay Divakaran
Ph.D., Rensselaer Polytechnic Institute, 1993
Principal Member Technical Staff, MERL Murray Hill

Ajay Divakaran (M'93) received the B.E. (with Hons.) degree in Electronics and Communication Engineering from the University of Jodhpur, Jodhpur, India, in 1985, and the M.S. and Ph.D. degrees from Rensselaer Polytechnic Institute, in 1988 and 1993 respectively. He worked as a research associate at the Indian Institute of Science before joining Iterated Systems Inc., Atlanta, GA in 1995. At Iterated Systems he worked on video-coding algorithms for video telephony and entertainment-quality video. In 1998 he joined MERL Murray Hill, where he has worked on video indexing and summarization with a view to MPEG-7 applications.

Alan Esenther
M.Sc., Boston University, 1993
Principal Technical Staff, MERL Cambridge Systems

Alan Esenther (M.S.C.E) enjoys Internet technologies, distributed software development, dynamic web programming, GUI work, and usability enhancements. He has taken an interest in finding simpler ways to accomplish software tasks -- ideally leveraging existing end-user resources. Recent work involves super-lightweight real-time distributed collaboration using unmodified web browsers. Previous work involved interactive page generation for mobile agents, an email gateway for offline surfing, transaction monitors, kernel-level volume management, and microprocessor development. Interests include rock climbing, adventure travel, and learning new technologies.

George Fang
B.Sc., California Institute of Technology, 1990
Member Technical Staff, MERL Murray Hill

George Fang received his B.Sc. degree from California Institute of Technology and became a member of Mitsubishi Electric Corporation's Kyoto Works in 1990. During the ten years working in Japan, he was a hardware engineer designing analog and digital consumer televisions for the American market and coordinated joint design efforts between Japan and the United States. He joined the Murray Hill Laboratory in February of 2001 with research objectives in wireless and network technologies.

James Fang
B.Sc., Columbia University, 1992
Member Technical Staff, MERL Murray Hill

James Fang received his B.Sc. from Columbia University in 1992 and did some graduate work there before joining Mitsubishi Electric in 1995. He worked on consumer televisions for three years before transferring to MERL Murray Hill in 1998. He is currently working on digital wireless communications.

William Freeman
Ph.D., Massachusetts Institute of Technology, 1992
Senior Research Scientist & Associate Director, MERL Cambridge Research

Freeman, William T. is a senior research scientist at MERL Cambridge Research, where he studies machine learning applied to computer vision and interactive applications of computer vision. As part of his doctoral work, he developed "steerable filters", a class of oriented filters useful in image processing and computer vision. In 1997 he received the Outstanding Paper prize at the Conference on Computer Vision and Pattern Recognition for work on applying bilinear models to separate "style and content". During 1987-88 he was a Foreign Expert at the Taiyuan University of Technology, Taiyuan, Shanxi, China

Sarah Frisken
Ph.D., Carnegie Mellon University, 1991
Senior Research Scientist, MERL Cambridge Research

Sarah Frisken has research interests in computer graphics, volume visualization and physically based modeling. She has led a team of researchers and students to build a knee arthroscopy simulator that incorporates high-quality rendering, haptic feedback and physical modeling to simulate interactions between surgical tools and a computer model derived from 3D MRI data. Her current work is with Adaptively Sampled Distance Fields, a general representation of shape for computer graphics, which provides intuitive manipulation, and editing of smooth surfaces with fine detail. Applications include electronic sculpting, volumetric effects for rendering, color gamut representation, path planning for CNC milling, and rapid prototyping.

Andrew Garland
Ph.D., Brandeis University, 2000
Visiting Scientist, MERL Cambridge Research

Andy Garland is interested in artificial intelligence issues relating to coordinating multiple agents. He have been recently been working on learning hierarchical task models from examples provided by a domain expert. Such models can be used, among other things, to represent shared beliefs about how an agent and a human can collaborate to accomplish a task. His thesis was on memory-based techniques for autonomous agents to learn how to better coordinate joint activities.

Daqing Gu
Ph.D., SUNY Stony Brook, 1999
Member Technical Staff, MERL Murray Hill

Daqing Gu received the BE degree from Tsinghua University, Beijing, China in 1987; the MS and Ph.D. degrees in electrical engineering from the State University of New York at Stony Brook, Stony Brook, NY in 1996 and 1999, respectively. He is currently a Member of Technical Staff at MERL Murray Hill. His research interests include mobility management and radio resource management for wireless communications, wireless networking.

Jainlin Guo
Ph.D., Windsor University, 1995
Principal Technical Staff, MERL Murray Hill

Jainlin Guo received his Ph.D. from Windsor University in 1995. He worked at Waterloo Maple for a year and a half as a software developer and then joined MERL Murray Hill in 1998. He has published seven research papers and his primary research interests include home networks, digital broadcasting, and wireless computing.

Bret Harsham
Massachusetts Institute of Technology
Principal Technical Staff, MERL Cambridge Systems

Bret Harsham joined MERL Cambridge Systems in January 2001 to pursue interests in speech user interfaces and speech-centric devices. Prior to joining MERL, Bret spent 3 1/2 years at Dragon Systems designing and implementing handheld and automotive speech products. Earlier, he was a principal architect of a Firewall and Virtual Pritvate Network product. Bret's other technical interests include distributed architectures, knowledge representation and language theory.

Jyhchau "Henry" Horng
Ph.D., Polytechnic University, 1998
Member Technical Staff, MERL Murray Hill

Henry Horng received the Ph.D. from Polytechnic University in 1998. He has worked as a research assistant at Polytechnic and as software developer and lecturer for Chung Cheng Institute of Technology, Taiwan. Henry joined MERL Murray Hill in 1999. He has published seven research papers and has one patent. His primary research interests include digital signal processing and communications.

Frederick J. Igo, Jr.
B.S., LeMoyne College
Senior Principal Technical Staff, MERL Cambridge Systems

Fred Igo's professional interests are in software development and its process. Starting at IPL Systems and continuing at MERL Cambridge Systems. Fred developed and managed teams developing wide-word microcode and tools as part of mainframe development projects. Two of his four mainframe projects won MELCO's President's Awards. Fred next moved into database and middleware development. During three projects he has worked with DCE, Distributed OLTP message queuing, MELCO's hardware database accelerator DIAPRISM (SPG) and OLAP databases like Oracle Express. Recently Fred has transitioned to Java development, developing applications and tools for MERL Cambridge Systems' mobile agent system Concordia.

Thouis "Ray" Jones
B.Sc., University of Utah, 1994
Member Technical Staff, MERL Cambridge Research

Ray Jones's research interests include real-time rendering methods, efficient data representations for graphics applications, and graphics modeling tools. He joined MERL Cambridge Reserach in 1999. His previous work included developing a 3D paint system and optimization work on the SunPHIGS structure walker. His current research is on high-quality real-time rendering systems.

Darren Leigh
Ph.D., Harvard University, 1998
Research Scientist, MERL Cambridge Research

Darren Leigh's research interests range from electronic hardware and communications to operating systems and signal processing. Before coming to MERL Cambridge Research he worked on the Harvard University/Planetary Society Billion-channel ExtraTerrestrial Assay (Project BETA), a search for microwave signals from extraterrestrial civilizations. Other previous research includes 3D microscopic scanning, desktop manufacturing and network architectures for multimedia. His current research includes the Personal Eyewitness and interfacing and applications of the M32R/D and Artificial Retina chips.

Neal Lesh
Ph.D., University of Washington, 1998
Research Scientist, MERL Cambridge Research

Neal Lesh has been studying techniques for improving human-computer, collaborative problem solving. Before coming to MERL Cambridge Research, he completed a thesis on scalable and adaptive goal recognition at the University of Washington, and worked as a postdoc with James Allen at the University of Rochester on the TRIPS collaborative planning project. At MERL Cambridge Research, he is working on systems for interactively solving hard combinatorial optimization problems, as well as on technology for building collaborative interface agents, called COLLAGEN. Recently he has begun working on techniques to allow people to share experiences using digital media.

Brenden Maher
M.S., Massachusetts Institute of Technology, 1998
Senior Technical Staff, MERL Cambridge Systems

Mr. Maher received his Masters in Media Arts and Sciences from the MIT Media Lab in 1998. His interest in computers began through developing Immersive Real-Time 3D Virtual Environments with the Boston Virtual Reality Group in 1992. Mr. Maher studied Tangible Interface design under Hiroshi Ishi and developed a new type of "Ambient Display" called the Bit Blower to provide information in the periphery of one's focus. He has worked in interactive media (CD-ROM's) and built a 3D data visualization system for analysis of Mechanical Engineering Data while at Molecular Geodesics, Inc.

Sergei Makar-Limanov
Ph.D., Stanford University, 1994
Principal Technical Staff, MERL Cambridge Systems

Sergei Makar-Limanov received his Bachelors degree from University of Chicago, and his PhD in Mathematics from Stanford University. After spending a few years in academia, Sergei decided to join the "real world" where he worked on such software projects as computer aided manufacturing for PTC corporation and supply management for Kewill PLC. His most recent work involved designing automated software scalability testing tools at Empirix Corporation. Sergei has joined MERL in May 2001 to work on the Concordia project.

Fernando Matsubara
M.Sc., University of Tokyo
Senior Principal Member Technical Staff, MERL Murray Hill

Fernando Matsubara received his B.Sc. degree from the Universidad Nacional Autonoma de Mexico and his ME degree from the University of Tokyo. He managed mathematical simulation projects at the Institute for Electrical Research (Mexico) for five years and joined Mitsubishi Electric in 1990, where he managed projects on distributed computing and networking, moving to the U.S. in 1995. His primary research areas are digital communications, network technology, and digital video.

David McDonald
Ph.D., Massachusetts Institute of Technology, 1980
Principal Technical Staff, MERL Cambridge Systems

David McDonald is a computational linguist. He pioneered the field of generation of natural language texts from a knowledge representation. After getting his degree in Artificial Intelligence from MIT he was on the Computer Science faculty at UMass Amherst where he directed the work of seven Ph.D. students in problems from tutoring to domain modeling to typography. After working at several natural language companies he has recently joined MERL as a member of the COLLAGEN project where works at the interface between the symbolic reasoning of Collagen and the probabilistic reasoning of speech understanding.

James McDonald
B.A., Rhode Island College, 1996
Systems Network Administrator, MERL Cambridge

Jimmy McDonald joined the MERL System Network Administration team in 2001. Before joining MERL, Jimmy held various systems administration positions in small and mid-sided companies.

Baback Moghaddam
Ph.D., Massachusetts Institute of Technology, 1997
Research Scientist, MERL Cambridge Research

Baback Moghaddam's research interests are in computational vision and image processing with special focus on probabilistic visual learning, statistical and neural network modeling and pattern recognition. Prior to coming to MERL Cambridge Reserach, Dr. Moghaddam was at the Vision and Modeling Group at the MIT Media Laboratory. As part of his dissertation, he developed an automatic face recognition system that won ARPA's 1996 "FERET" face recognition competition. Past research includes fractal image compression, segmentation and analysis of SAR and IR imagery as well as designing a zero-gravity experiment for laser annealing of amorphous silicon flown aboard the U.S. space shuttle in 1990.

Yves-Paul Nakache
M.Sc., E.S.I.E.E., 2000
Member Technical Staff, MERL Murray Hill

Yves-Paul Nakache received a French Engineering diploma equivalent to M.Sc. degree in Electrical Engineering in 2000 from E.S.I.E.E. (Ecole Supérieure d'Ingénieurs en Electrotechnique et Electronique), Paris, France. He joined MERL Murray Hill in 2000, where he is currently a Member of the Technical Staff. He works on Interference Cancellation and 3G CDMA systems. His current interests are in speech processing and wireless communications.

Philip Orlik
Ph.D., SUNY Stony Brook, 1999
Member Technical Staff, MERL Murray Hill

Philip Orlik received the B.E. degree in 1994 and the M.S. degree in 1997 both from the State University of New York at Stony Brook. In 1999 he earned his Ph. D. in electrical engineering also from SUNY Stony Brook. He joined MERL Murray Hill in August 2000, and is currently a member of technical staff. His research interests include wireless and optical communications, networking, queuing theory, and analytical modeling.

Kadir Peker
Ph.D., New Jersey Institute of Technology, 2001
Member Technical Staff, MERL Murray Hill

Kadir A. Peker received the B.S. degree from Bilkent University, Turkey in 1993, the M.S. degree from Rutgers University in 1996, and the Ph.D. degree from New Jersey Institute of Technology in 2001, all in Electrical Engineering. His Ph.D. dissertation is on content based video indexing and summarization using motion activity. He worked at MERL Murray Hill as an intern for more than a year, and joined as a member of technical staff in 2000. He has contributed to MPEG-7, published conference papers, submitted journal papers and filed patents on his Ph.D. topic. His current research interests include video indexing and summarization, home networking and multimedia networks.

Ron Perry
B.Sc., Bucknell University, 1981
Research Scientist, MERL Cambridge Research

Ron Perry joined MERL Cambridge Research as a Research Scientist in 1998. Prior to that, he was a consulting engineer at DEC developing a three-dimensional rendering ASIC called Neon. Ron has consulted for many companies including Kodak, Adobe, Quark, and Apple over the last 18 years, developing software and hardware products in the areas of computer graphics, imaging, color, and desktop publishing. Some key product developments include the color engines for QuarkXPress, Adobe PhotoShop, Adobe Illustrator, and Windows 95/98 as well as the Atex Display and Pagination system which is used by most major metropolitan newspapers in the world to paginate, display, and print their publications.

Hanspeter Pfister
Ph.D., State University of New York at Stony Brook, 1996
Research Scientist, MERL Cambridge Research

Hanspeter Pfister is a Research Scientist at MERL. He is the chief architect of VolumePro, Mitsubishi Electric's real-time volume rendering hardware for PCs. His research interests include computer graphics, scientific visualization, and computer architecture. Hanspeter Pfister received his Ph.D. in Computer Science in 1996 from the State University of New York at Stony Brook. In his doctoral research he developed Cube-4, a scalable architecture for real-time volume rendering. He received his M.S. in Electrical Engineering from the Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, in 1991. He is a member of the ACM, ACM SIGGRAPH, IEEE, the IEEE Computer Society, and the Eurographics Association.

Erik Piip
Manager, Computer Network Services, MERL Cambridge

Erik is the manager of the Computer Network Services Group. The group supports MERL's computing and network infrastructure, and end-users. Erik is responsible for identifying MERL wide strategic and tactical enhancements. In Part lives, Erik worked for Digital Equipment in multiple roles; service delivery, support, fault management and analysis, and server management design. Other interests include using and building large telescopes and instrumentation and Amateur Radio.

Fatih Porikli
Ph.D., Polytechnic University, 2001
Member Technical Staff, MERL Murray Hill

Fatih Porikli received the B.S. degree from the Bilkent University, Turkey in 1992, and the M.S. and Ph.D. degrees in electrical engineering from the Polytechnic University, Brooklyn, NY in 1995 and 2001, respectively. From 1999 to 2000, he worked for Hughes Research Labs, Malibu, CA. He joined to Mitsubishi Electric Research Labs at Murray Hill in August 2000. His research interests are in the areas of video processing, computer vision, 3-D depth estimation, aerial image processing, texture segmentation, robust optimization, network traffic, and digital signal filtering.

Stanley Pozerski
BA Computer Systems, Daniel Webster College 1987
Systems Network Administrator, MERL Cambridge

Stan's interests have followed the application of computers to a variety of manufacturing tasks including using PDP-11's to demonstrate control of multiple reactor chemical processes, using personal computers for production testing, manufacturing chemicals and controlling multi-axis rotary assembly machines. More recently, Stan has been Systems Administrator of a CIM system for a semiconductor facility performing shop floor scheduling, data collection, and process monitoring. Currently, Stan supports Windows and Linux clients and servers, networking, and the wide variety of PC applications used at MERL Cambridge Research.

Bhiksha Raj
Ph.D., Carnegie Mellon University, 2000
Research Scientist, MERL Cambridge Research

Dr. Bhiksha Raj joined CRL as a Staff Scientist. He completed his Ph.D. from Carnegie Mellon University (CMU) in May 2000. He way employed at Compaq's Cambridge Research Labs from June 2000 to May 2001, when he joined MERL. Dr. Raj works mainly on algorithmic aspects of speech recognition, with special emphasis on improving the robustness of speech recognition systems to environmental noise. He has over twenty-five conference and journal publications and is currently in the process of publishing a book on missing-feature methods for noise-robust speech recognition.

Ramesh Raskar
Ph.D., University of North Carolina at Chapel Hill, 2001
Research Scientist, MERL Cambridge Research

Ramesh Raskar joined MERL Cambridge Research as a Research Scientist in 2000. Prior to that, he was at the Office of the Future group at UNC's Computer Graphics lab. As part of his dissertation, he developed a framework for projector based 3D graphics by treating a projector as the dual of a camera. Current work includes topics from non-photorealistic rendering, computer vision and intelligent user interfaces. He is a member of the ACM.

Charles Rich
Ph.D., Massachusetts Institute of Technology, 1980
Senior Research Scientist, MERL Cambridge Research

The thread connecting all of Dr. Rich's research has been to make interacting with a computer more like interacting with a person. As a founder and director of the Programmer's Apprentice project at the MIT Artificial Intelligence Laboratory in the 1980s, he pioneered research on intelligent assistants for software engineers. Dr. Rich joined MERL Cambridge Research in 1991 as a founding member of the research staff. For the past several years, he has been working on a technology for building collaborative interface agents, called COLLAGEN, which is based on a theory of human collaborative discourse. Dr. Rich is a Fellow and past Councilor of the American Association for Artificial Intelligence.

David Rudolph
M.S., University of Illinois, 1989
Principal Technical Staff, MERL Cambridge Systems

David Rudolph has been at CSL for 10 years. During this time he has contributed to several systems software projects, including the gm80 simulator. His last 4 ½ years have been spent developing the "Network Replication" project, which is now being successfully marketed by Veritas Software Corporation as SRVM (Storage Replicator for Volume Manager). This project was the 1998 winner of the Corporate R&D GM's Award for Excellence. Before joining CSL, David spent three years at Data General, interrupted by a two-year stint at the University of Illinois, where his research interests were system software and performance analysis for massively parallel architectures.

Kathy Ryall
Ph.D., Harvard University, 1997
Principal Technical Staff, MERL Cambridge Systems

Kathy Ryall received the B.A. degree with a double major in Mathematics/Computer Science and Psychology from Wesleyan University, and the M.S. and Ph.D. degrees in Computer Science from Harvard University. Prior to joining ITA/HSL in July 2000, Kathy served as an Assistant Professor of Computer Science at the University of Virginia for three years. Kathy's primary technical interest is the design and implementation of systems in which the interface acts as a medium for people and computers to work together on solving problems, rather than as a means for people to control computers. This technical interest has resulted in the development of systems for a variety of tasks in the computer graphics and graphic design domains.

Zafer Sahinoglu
M.S., NJIT
Technical Staff, MERL Murray Hill

Zafer Sahinoglu received his BS in Electrical and Electronics Engineering from Gazi University, Ankara Turkey and his MS in Biomedical Engineering from NJIT, Newark, NJ. He is currently pursuing his PhD in Internet Traffic Engineering at NJIT. He worked at AT&T Shannon Labs in the Advanced Local Network Technology Group in 1999. He joined MERL in March 2001. His area of research includes DSL modem technologies and networks, dynamic resource allocation for multimedia applications, QoS provisioning, and traffic self-similarity. He has published several conference papers submitted two transactions papers and filed three patent applications.

Bent Schmidt-Nielsen
B.S. University of California at San Diego, 1971
Senior Principal Technical Staff, MERL Cambridge Systems

Bent Schmidt-Nielsen has seven years of experience at Dragon Systems in applying speech recognition to useful products. Here at MERL he is paying a lot of attention to making speech interfaces robust and usable. Bent has very broad interests in science and technolgy. Among many other activities he has taught genetics at the University of Massachusetts at Boston and he has been a leader in the development of an easy to use mass market database.

Derek Schwenke
M.S., Worcester Polytechnic Institute
Principal Technical Staff, MERL Cambridge Systems

Derek Schwenke received his B.S.E.E. from Tulane and M.S.C.S. from Worcester Polytechnic Institute. His work areas have included the computer software and hardware fields. He has worked with image processing and satellite communications systems at Raytheon Corp. in Marlboro, MA. Derek began his work at MERL Cambridge Systems in 1988 with the design and simulation of M80 and PXB1 CPUs, and progressed to software development with the OSF-DCE/Encina system. He co-developed the OMQ message queuing system for MELCO. In the last three years Derek has worked extensively with Open Community virtual reality system. Most recently he's helped develop the Internet Sharing and Transfer Protocol (ISTP).

Mark Shabelman
B.S., Chernivtsi State University, 1996
System Network Administrator, MERL Murray Hill

Mark Shabelman received his B.S. in Computer Science from Chernivtsi State University in 1996. In 2000 he graduated as an honor student from CHUBB University, Parsippany, NJ, specializing in computer programming. He joined the MHL in April 2000. Previously, Mark has worked as a System Administrator at Impuls LLC, Inc, Chernivtsi, Ukraine.

Chia Shen
Ph.D., University of Massachusetts, 1992
Senior Research Scientist, MERL Cambridge Research

Chia Shen's current research is in distributed real-time and multimedia systems. She is particularly interested in the non-traditional use of standard high-speed networks, such as ATM, for distributed industrial control applications and distributed multimedia environments. Her work involves the design of middleware algorithms and protocols. In her Ph.D. research on the Spring kernel, she worked on issues in multiprocessor real-time operating systems including on-line scheduling, dispatching, resource reclamation, and predictable synchronization mechanisms.

Samuel Shipman
M.Sc., Carnegie Mellon University, 1985
Principal Technical Staff, MERL Cambridge Systems

Sam Shipman received the M.S. degree in Computer Science from Carnegie Mellon University and the B.S. from UNC-Wilmington. His technical interests and background are in real-time and distributed operating systems research and development. At MERL Cambridge Systems, he has worked on the Network Replication and Open Community projects, and on smaller efforts related to MPEG-7, interactive surroundings, and fingerprint recognition.

Candace Sidner
Ph.D., Massachusetts Institute of Technology, 1979
Senior Research Scientist, MERL Cambridge Research

Candy Sidner is an expert in user interfaces, especially those involving speech and natural language understanding, and human and machine collaboration. Before coming to MERL, she had been a research staff member at Bolt Beranek Newman, Digital Equipment Corp., and Lotus Development Corp., and a visiting scientist at Harvard University. She is currently working on how to apply speech understanding technology to collaborative interface agents in the COLLAGEN project. Dr. Sidner was Chair of the 2001 International Conference on Intelligent User Interfaces and is a past President of the Association for Computational Linguistics. She is also a Fellow and past Councilor of the American Association for Artificial Intelligence.

Huifang Sun
Ph.D., University of Ottawa, 1986
Deputy Director, MERL Murray Hill

Huifang Sun received his B.Sc. degree from Harbin Engineering Institute, Harbin, China in 1967, and his Ph.D. from University of Ottawa, Canada, in 1986. He was an Associate Professor at Fairleigh Dickinson University before moving to Sarnoff Research Laboratory in 1990, where he was Technology Leader of Digital Video Communication. He joined MERL Murray Hill in 1995. His research interests include digital video/image compression and digital communication.

Jeroen van Baar
M.Sc., Delft University of Technology, 1998
Member Technical Staff, MERL Cambridge Research

Jeroen van Baar joined MERL Cambridge Research in 1999 as Member of Technical Staff. His interests are in the broad fields of computer graphics, scientific visualization and user interfaces. Jeroen is currently working on a number of different projects including EWA Surface / Volume Splatting, 3D Images, Smart Projectors, and I-Chamber. Jeroen has been a Teaching Assistant for the Introduction to and Advanced Computer Graphics courses at the Harvard University Extension School for the past two years.

Anthony Vetro
Ph.D., Polytechnic University, 2001
Principal Member Technical Staff, MERL Murray Hill

Anthony Vetro received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from Polytechnic University, Brooklyn, NY. He joined MERL in 1996, where he is currently a Principal Member of the Technical Staff. At MERL, he worked on algorithms for down-conversion decoding of compressed video signals, which were implemented into MELCO's second-generation HDTV receiver chip. His current research interests are in the area of multimedia coding and transmission, with emphasis on content scaling and optimal resource allocation. He has also been an active participant in MPEG standards for several years.

Paul Viola
Ph.D., Massachusetts Institute of Technology, 1995
Research Scientist, MERL Cambridge Research

Before moving to MERL Paul Viola was an Associate Professor of Computer Science and Engineering at the Massachusetts Institute of Technology. He also spent two years as a visiting scientist in the Computational Neurobiology of the Salk Institute in San Diego. Paul has a broad background in advanced computational techniques, publishing in the fields of computer vision, neurobiological vision, medical imaging, mobile robotics, machine learning, and automated drug design. Paul was a recipient of a National Science Foundation Career award in 1998. He has worked on research and development with a number of companies including: Compaq, IBM Research, Arris Pharmaceuticals and Intarka.

Kent Wittenburg
Ph.D., University of Texas at Austin, 1986
Senior Principal Technical Staff, MERL Cambridge Systems

Kent Wittenburg joined MERL in July 2001. He is interested in novel human-computer interaction technologies and devices. Past work has included consumer-based information visualization, visual language parsing, and natural language understanding. He comes to MERL from Verizon Laboratories, where he led the Advanced Interactive Internet Technologies group. He has also worked on the research staffs of Bellcore and MCC. He holds a Ph.D. from University of Texas at Austin received in 1986.

Joseph Woelfel
M.S., Rutgers University, 1992
Principal Technical Staff, MERL Cambridge Systems

Before joining MERL in February 2001, Joe worked at Dragon Systems, where he led small teams developing an extensible voice architecture. In the years before that, Joe worked on the development of a statistical process control software package at GE-Fanuc. Joe earned a B.S. in Physics from SUNY Albany, and an M.S. in Communication and Information Science from Rutgers University. As Project Engineer for the Surveillance product, he will work on MERL's Interactive Surroundings Initiative.

Peter Wolf
B.S., Yale University, 1983
Senior Principal Technical Staff, MERL Cambridge Systems

Peter is an expert in Speech Technologies and a broad range of Software Engineering tools and practices. While his role is often that of a technical expert and principal engineer, his main interest is the definition and creation of new products and services, made possible by new technologies. Peter has a 20 year history of leading teams of engineers from initial product concepts, to the production of shipping products. Some of the successful shipping projects where Peter was a principal member include: Spoken Controls for Automobiles, Embedded Speech Recognition Toolkit, Digital Video Editor/Database/Network Software, 3d Graphics CAD System.

David Wong
Ph.D., University of Connecticut, 1991
Deputy Director, MERL Cambridge Systems

David Wong is an expert in mobile agent technology and the primary evangelist for Mitsubishi's mobile agents initiative. His background also includes substantial work in transactional message queuing systems and distributed transaction processing. Prior to joining MERL in 1994, David worked on the advanced development and performance analysis of transaction processing systems at Compaq. He has also taught and conducted research at Brown University and the University of Connecticut. David holds a B.S. from Brown University and a Ph.D. in computer science from the University of Connecticut.

Chris Wren
Ph.D., Massachusetts Institute of Technology, 1999
Research Scientist, MERL Cambridge Research

Christopher Wren's research area is Perception for Human-Computer Interaction. While Chris' recent work has focused on using computer vision techniques to create systems that are visually aware of the user, his current interests also extend to include audio processing and other sensing modalities. Prior to coming to MERL, he was at the Vision & Modeling Group at the MIT Media Laboratory. As part of his dissertation work, he developed a system for combining physical models with visual evidence in real time to recover subtle models of human motion.

Jonathan Yedidia
Ph.D., Princeton University, 1990
Member Technical Staff, MERL Cambridge Research

Jonathan Yedidia's graduate work at Princeton and post-doctoral work at Harvard's Society of Fellows focused on theoretical condensed-matter physics, particularly the statistical mechanics of systems with "quenched" disorder. He was a professional chess player and teacher from 1993 to 1997. He then joined the internet startup company Viaweb, where he worked on a shopping search engine that has since become shopping.yahoo.com. Dr. Yedidia is interested in the application of statistical methods to inference and learning.

William Yerazunis
Ph.D., Rensselaer Polytechnic Institute, 1987
Research Scientist, MERL Cambridge Research

William Yerazunis has worked in a number of fields including: optics and signal processing (for General Electric's jet engine manufacturing); computer graphics (at Rensselaer's Center for Interactive Computer Graphics); artificial intelligence and parallel symbolic computation (for DEC's OPS5, XCON, and the successor products such as RuleWorks); radioastronomy and exobiology (at Harvard University), and transplant immunology (for the American Red Cross). He holds 15 U.S. patents.

Charles You
M.Sc., University of Buffalo, 1999
Member Technical Staff, MERL Murray Hill

Charles You received the B.S. and M.S in electrical engineering from the University of Buffalo, New York and the Columbia University, New York in 1995 and 1999, respectively. He has been employed by MERL Merray Hill since February 2000, working in the area of DSPs and wireless mobile systems. His main research interests are in the general areas of DSPs and communications.

Fangfang Zhang
M.S., Brandies University, 2000
Technical Staff, MERL Cambridge Systems

Fangfang Zhang received her Master in Software Engineering from Brandies University. She joined MERL-CSL in Jan 2001. She is currently working on product building and testing for Concordia development team. Prior to joining CSL, Fangfang worked shortly at CMGI, as a member of web-dialup service development team. She also worked at Scriptgen Pharmaceuticals, Inc., where she designed relational database and implemented a number of software to assist biophysics research.

Jinyun Zhang
Ph.D., University of Ottawa
Senior Principal Member Technical Staff, MERL Murray Hill

Jinyun received her Ph.D. in Electrical Engineering from the University of Ottawa, where she was also a Visiting Scholar. Prior to this she was a teacher/lecturer at Tsinghua University, Beijing, China. Jinyun worked for Nortel Networks for the past ten years where she held engineering and most recently project management positions in the areas of Advanced Wireless Technology Development, Wireless Networks and Optical Networks. She has a broad technical background, specializing in system design for wireless communications and DWDM optical networks. Jinyun joined the Murray Hill Laboratory in June of 2001.

Remo Ziegler
M.Sc., Swiss Federal Institute of Technology ETH, 2001
Member Technical Staff, MERL Cambridge Research

During Remo's studies of Computer Science, he specialized in computer graphics and computer vision. In term papers and his master thesis, he applied the knowledge in biomedical problems or virtual surgery. He gained most of his experience in computer graphics while preparing his master thesis, where he worked on a real time simulation of natural forces of skin and tissue as they resist incision by a scalpel. Remo joined MERL Cambridge Research in 2001 and is currently working on a program generating a 3D polygonal model out of several projections.

Recent Major Publications

The following lists the major publications by members of the MERL staff over the past 2 years. A publication is considered major if it appeared in a refereed journal or a refereed conference proceedings. For completeness, the list includes a number of publications that have been accepted for publication in the near future.

2001

Frisken, Sarah F., Perry, Ronald N., "Computing 3D Geometry Directly from Range Images." To appear in *SIGGRAPH 2001 Conference Abstracts and Applications*. (TR2001-10)

Perry, Ronald N., "A New Method For Numerical Constrained Optimization." To appear in *SIGGRAPH 2001 Conference Abstracts and Applications*. (TR2001-14)

Bell, Matt and Freeman, William T., "Learning local evidence for shading and reflectance", *Proc. Intl. Conf. Computer Vision* 2001, July, 2001, (TR2001-04)

Pope, Jackson, Frisken, Sarah F., Perry, Ronald N. "Dynamic Meshing Using Adaptively Sampled Distance Fields." To appear in *SIGGRAPH 2001 Conference Abstracts and Applications*. (TR2001-13)

Frisken, Sarah F., Perry, Ronald N., "A Computationally Efficient Framework for Modeling Soft Body Impact." To appear in *SIGGRAPH 2001 Conference Abstracts and Applications*. (TR2001-11)

Yedidia, Jonathan S., Freeman, William T., and Weiss, Yair, "Understanding Belief Propagation and its Generalizations," to appear in IJCAI-2001 Book of Distinguished Papers, (TR2001-27)

Yedidia, Jonathan S., and Bouchaud, Jean-Philippe, "Renormalization Group Approach to Error-Correcting Codes," delivered at the ICTP Workshop on Statistical Physics and Capacity-Approaching Codes, Trieste 2001, (TR2001-19)

Zwicker, Mathius, Pfister, Hanspeter Pfister, van Baar, Jeron, Gross, Matt, "EWA Volume Splatting." To appear in *Proceedings of Visualization 2001*.

Zwicker, Mathius, Pfister, Hanspeter, van Baar, Jeron, Gross, Matt, "Surface Splatting." To appear in *Proceedings of SIGGRAPH 2001*. (TR2001-21)

Dietz, Paul H., Yerazunis, William S., "Real-Time Audio Buffering for Telephone Applications", to appear in *UIST 2001 Conference Proceedings,* November 2001.

Dietz, Paul H., Leigh, Darren, "DiamondTouch: A Multi-User Touch Technology", to appear in *UIST 2001 Conference Proceedings*, November 2001.

Moghaddam, Baback., Nastar, Chahab., and Pentland, Alex. "A Bayesian Similarity Measure for Deformable Image Matching," *Image & Vision Computing*, Elsevier, Vol. 19, pps. 235-244, 2001.

Moghaddam, Baback, and Yang Ming-Hsuan, "Sex with Support Vector Machines," *Advances in Neural Information Processing Systems 13*, MIT Press, 2001.

Moghaddam, Baback, Biermann, Henning, and Margaritis, Dimitris. "Regions-of-Interest and Spatial Layout for Content-Based Image Retrieval," *Multimedia Tools & Applications*, Kluwer, Vol. 14, No. 2, pps. 199-208, 2001.

Divakaran, Ajay, R. Regunathan, and Peker, Kadir, "Video summarization with motion descriptors," *Journal of Electronic Imaging*, vol. 10, no. 4, October 2001

Zwicker, Matthias, Pfister, Hanspeter, van Baar, Jeroen, Gross, Markus, "EWA Volume Splatting", to appear in *IEEE Visualization 2001 Conference Proceedings*, October 2001

Garland, Andrew, Ryall, Kathy, Rich, Charles, "Learning Hierarchical Task Models by Defining and Refining Examples" *First International Conference on Knowledge Capture* October, 2001, (TR2001-26)

Moghaddam, Baback, Zhou, Xiang, and Huang, Thomas S., "ICA-based Probabilistic Local Appearance Models," *International Conference on Image Processing* (ICIP'01), Thessaloniki, Greece, October 8-10th, 2001.

Y. Wu, Vetro, Anthony, Sun, Huifang, and S.Y. Kung, "Intelligent multi-hop video communications," *Proc. IEEE Pacific-Rim Conference on Multimedia*, Beijing, China, October 2001.

X. Sun, Divakaran, Ajay, B.S. Manjunath, "A motion activity descriptor and its extraction in the compressed domain" *Proc. IEEE Pacific-Rim Conference on Multimedia*, Beijing, China, October 2001.

Vetro, Anthony,Y. Wang and Sun, Huifang, "Rate-distortion optimized video coding considering frameskip," *Proc. IEEE Int'l Conf. on Image Processing*, Thessaloniki, Greece October 2001.

Peker, Kadir, Divakaran Ajay, and Sun,Huifang, "Constant pace skimming and temporal sub-sampling of video using motion activity," *Proc. IEEE Int'l Conf. on Image Processing*, Thessaloniki, Greece October 2001.

Porikli, Faith"Accurate detection of edge orientation for color and multi spectral imagery," *Proc. IEEE Int'l Conf. on Image Processing*, Thessaloniki, Greece October 2001

Porikli, Faithand Y. Wang, "An Unsupervised Multi-resolution Object Extraction Algorithm Using Video-Cubes" *Proc. IEEE Int'l Conf. on Image Processing*, Thessaloniki, Greece October 2001.

Casey, Michael, "Reduced-Rank Spectra and Entropic Priors as Consistent and Reliable Cues for General Sound Recognition", *Proceedings of the Workshop on Consistent & Reliable Acoustic Cues for Sound Analysis*, September, 2001.

Moghaddam, Baback, Tian, Qi, Lesh, Neal., Shen, Chia., and Huang, Thomas S., "Visualization and Layout for Personal Photo Libraries", *Content-Based Multimedia Indexing* (CBMI'01), Brescia, Italy, September 18-20th, 2001.

Divakaran, Ajay "An overview of MPEG-7 descriptors and their applications," *Proc. Int'l Conf. On Computer Analysis of Images and Patterns*, Warsaw, Poland, September 2001.

Moghaddam, Baback, Tian, Qi., Lesh, Neal., Shen, Chia., and Huang, Thomas S., "Visualization and Layout for Personal Photo Libraries", *Content-Based Multimedia Indexing (CBMI'01)*, Brescia, Italy, September 18-20th, 2001.

Moghaddam, Baback, Tian, Qi., Lesh, Neal., Shen, Chia., and Huang, Thomas S., "Visualization and Layout for Personal Photo Libraries", *Content-Based Multimedia Indexing* (CBMI'01), Brescia, Italy, September 18-20th, 2001.

Porikli, Faith"Object segmentation of color video sequences," *Proc. Int'l Conf. On Computer Analysis of Images and Pattern*, Warsaw, Poland, September 2001.

Perry, Ronald N.,Frisken F. Sarah, "Kizamu: A System For Sculpting Digital Characters", to appear in *SIGGRAPH 2001 Conference Proceedings*, August 2001 (TR2001-08)

Rich, Chuck., Lesh, Neal., Sidner, Candy., "Human-Computer Collaboration for Universal Access," *First Int. Conf. on Universal Access in Human-Computer Interaction*, New Orleans, LA, August 2001 (Invited Paper).

Zwicker, Matthias, Pfister, Hanspeter, van Baar, Jeroen, Gross, Markus, "Surface Splatting", to appear in *Siggraph 2001 Conference Proceedings*, August 2001

Casey, Michael, "General Sound Similarity and Recognition: Musical Applications of MPEG-7 Audio", *Organized Sound*, August 2001.

P. Xu, S.F. Chang, Divakaran, Ajay, Vetro, Anthony, and Sun, Huifang "Algorithm and system for high-level structure analysis and event detection in soccer video," *Proc. IEEE Int'l Conf on Multimedia and Expo*, Tokyo, Japan, August 2001.

Peker, Kadir and Divakaran, Ajay "A novel pair-wise comparison based analytical framework for automatic measurement of intensity of motion activity of video *Proc. IEEE Int'l Conf on Multimedia and Expo*, Tokyo, Japan, August 2001.

Vetro, Anthony,Y. Wang and Sun, Huifang "Estimating distortion for coded and non-coded frames for frameskip-optimized video coding," *" Proc. IEEE Int'l Conf on Multimedia and Expo*, Tokyo, Japan, August 2001.

Raskar, Ramesh, "Hardware Support for Non-photorealistic Rendering", *Siggraph Eurographics Graphics Hardware Workshop*, Los Angeles, August 2001 TR2001-23

Cassell, Justine, Nakano, Yukiko I., Bickmore, Timothy W., Sidner, Candace L. and Rich, Charles. "Non-Verbal Cues for Discourse Structure," to appear in the *Proceedings of the 39th meeting of the Association for Computational Linguistics*, July 2001, (TR2001-24)

Lesh, Neal, Rich, Charles and Sidner, Candace L., "Collaborating with Focused and Unfocused Users under Imperfect Communication," to appear in the *Proceedings of the 8th Int. Conf. on User Modelling*, July 2001 (TR2000-30)

Rich, Charles, Sidner, Candace L., Lesh, Neal., "Collagen: Applying. Collaborative Discourse Theory to Human-Computer Interaction," to appear in *Artificial Intelligence Magazine*, Special Issue on Interface Agents, 2001, (TR 2000-38)

Horng, Henry, and Nakache,Yves-Paul, "Using diversity combining for Adaptive interference cancellation in CDMA systems", *Proc. Int'l Symp. on Signals, Systems and Electronics*, Tokyo, Japan, July 2001.

Chen, Horng, Henry, J. Bao, and H. Kobayashi, "A joint channel estimation and signal detection algorithm for OFDM system", *Proc. Int'l Symp. on Signals, Systems and Electronics*, Tokyo, Japan, July 2001

Efros, Alexi A. and Freeman, William T., "Image Quilting for Texture Synthesis and Transfer", *Trans. ACM SIGGRAPH* 2001, August, 2001, (TR2001-17)

Jeannin S.,and Divakaran, Ajay "MPEG-7 visual motion descriptors," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 6, June 2001.

Raskar, Ramesh., Welch, Greg., Low, Kok-lim., Bandyopadhyay, Deepak., "Shader Lamps: Animating Real Objects with Image Based Illumination", Eurographics Workshop on Rendering, London, June 2001 (TR2001-21)

X. Zhang, Vetro, Anthony, Sun, Huifang, and Y-Q. Shi, "Robust decoding for reduced error propagation of DC/AC prediction errors," *Proc. 2nd Workshop and Exhibition on MPEG-4*, San Jose, CA, June 2001.

Vetro, Anthony, Divakaran, Ajay and Sun, Huifang, "Providing multimedia services to a diverse set of consumer devices," *Proc. Int'l Conf. On Consumer Electronics*, Los Angeles, CA, June 2001.

You, Charles, and Horng, Henry, "Optimum frequency and symbol synchronization for OFDM systems," *Proc. Int'l Conf. On Consumer Electronics*, Los Angeles, CA, June 2001.

Horng ,Henry and Derand, Thierry, "Interference cancellation in CDMA downlink supporting variable rate services", *Proc. Int'l Conf. On Consumer Electronics*, Los Angeles, CA, June 2001.

Wang, Haining., Shen, Chia.,, and Shin., Kang., "Adaptive-Weighted Packet Scheduling for Premium Service", ICC2001 *(IEEE International Conference on Communications),* June 11-14, 2001.

Casey, Michael, "General Sound Similarity and Sound Recognition Tools", in B.S. Manjunath, P. Salembier and T. Sikora, (Eds), *Introduction to MPEG-7: Multimedia Content Description Language*, J. Wiley, 2001

Casey, Michael "MPEG-7 Sound Recognition Tools", IEEE Transaction on Circuits and Systems Video Technology, special issue on MPEG-7, *IEEE*, June 2001.

Rich, Charles, Sidner, Candace L., Lesh, Neal., "Collagen: Applying Collaborative Discourse Theory to Human-Computer Interaction," to appear in *Artificial Intelligence Magazine*, Special Issue on Interface Agents, 2001, (TR 2000-38)

Rickel, J., Lesh, Neal., Rich, Chuck., Sidner, Candy., Gertner, A., "Building a Bridge between Intelligent Tutoring and Collaborative Dialogue Systems," *Proc 10th Int. Conf. on Artificial Intelligence in Education*, San Antonio, TX, May 2001, pp. 592-594. (Outstanding Paper Award)

Vetro Anthony, and Sun, Huifang, "Media conversions to support mobile users," *Proc. IEEE Canadian Conf. on Electrical and Computer Engineering*, Toronto, Canada, May 2001.

Vetro Anthony, and Sun, Huifang, "Encoding and transcoding multiple video objects with variable temporal resolution," *Proc. Int'l Symposium Circuits and Systems*, Sydney, Australia, May 2001.

Kuhn, P., Suzuki , T., Vetro, Anthony, "MPEG-7 transcoding hints for reduced complexity and improved quality," *Proc. Int'l Packet Video Workshop*, Kyongju, Korea, April 2001.

Shen, Chia., Lesh, Neal., Moghaddam, Baback., Beardsley, Paul.,and Bardsley, Ryan S., "Personal Digital Historian: User Interface Design", *CHI-2001*, April 1-5, 2001.

Vetro, Anthony and Sun, Huifang, "An overview of MPEG-4 object-based encoding algorithms," *Proc. Int'l Conf. on Information Technology: Coding and Computing*, Las Vegas, NV, April 2001.

Vetro, Anthony, Sun Huifang, and Y. Wang, "Object-based transcoding for adaptable video content delivery," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 3, March 2001.

Aish, Robert, Frankel, James L., Frazer, John H., Patera, Anthony T., and Marks, Joe, "Computational Construction Kits for Geometric Modeling and Design (Panel)," *Proceedings of the 2001 Symposium on Interactive 3D Graphics, Research Triangle Park,* North Carolina, March 2001, (TR2001-02)

Marks, Joe, Freeman, William, and Leitner, Henry, "Teaching Applied Computing without Programming: A Case-Based Introductory Course for General Education," *Proceedings of the SIGCSE 2001 Technical Symposium on Computer Science Education,* Charlotte, North Carolina, February 2001, TR2001-06.

Davies, Jim R., Gertner, Abigail, Lesh, Neal, Rich, Charles, Rickel, Jeff and Sidner, Candace L., "Incorporating Tutorial Strategies into an Intelligent Assistant," in the *Proceedings of the Fifth Int. Conf. on Intelligent User Interfaces,* Santa Fe, NM, January 2001, (TR2000-30)

Peker, Kadir, and Divakaran, Ajay, "Automatic measurement of intensity of motion activity of video segments," *Proc. SPIE Conf. on Storage and Retrieval from Multimedia Databases*, San Jose, CA, January 2001.

Divakaran ,Ajay,and Peker, Kadir, "Video summarization using motion descriptors," *Proc. SPIE Conf. on Storage and Retrieval from Multimedia Databases*, San Jose, CA, January 2001.

2000

Vetro, Anthony, Divakaran, Ajay, Sun, Huifang, and Poon, Tommy, "Transcoding system based on MPEG-7 meta-data," *Proc. IEEE Pacific-Rim Conf. on Multimedia*, Sydney, Australia, December 2000.

Horng, Henry and J. Bao, "An adaptive diversity combining digital receiver for multiple multiplexing schemes", *Proc. Int'l Symposium on Intelligent Signal Processing and Communication Systems*, Honolulu, HI, November 2000.

Moghaddam, Baback., Jebara, Tony., and Pentland, Alex. "Bayesian Face Recognition," *Pattern Recognition*, Vol. 33, No. 11, pps. 1771-1782, November, 2000.

Shen, Chia., and Mizunuma, Ichiro., "RT-CRM: Real-Time Channel-Based Reflective Memory", *IEEE Transactions on Computers*, Vol 49, No 11. November 2000.

Chung, Sun-Tae., Gonzalez, Oscar., Ramamritham, Krithi., Shen,Chia., "CReMeS: A CORBA Compliant Reflective Memory based Real-time Communication Service", *IEEE 2000 Real-Time Systems Symposium*, November 27-30, 2000 (TR2000-18)

Bober, M., Asai, K., and Divakaran, Ajay "MPEG-4/7 based internet video and still image browsing system," *Proc. SPIE Conference on Multimedia Systems and Applications*, Boston, MA, November 2000.

Strohecker, Carol, Slaughter, Adrienne, "Zyklodeon: A Software Construction Kit Modeling Cyclic Timing Patterns" *Proceedings of ACM Multimedia*, November 2000 (TR2000-29)

Gu, D., and Bao, J., "Performance evaluation of a high-speed data transmission scheme with modulo-type preequalization," *Proc. 34th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, October 2000.

Freeman, William T., Haddon, John A., and Pasztor, Egon C., "Learning motion analysis", in *Statistical Theories of the Brain*, MIT Press, edited by R. Rao, B. Olshausen and M. Lewicki, 2001, (TR2000-32)

Freeman, William T., book review of "Biometrics: personal identification in a networked society", in *Pattern Analysis and Applications* 4:67-68, April, 2001, (TR2000-07)

Freeman, William T., Pasztor, Egon C., and Carmichael, Owen T., "Learning low-level vision", *Int. J. of Computer Vision* 40(1), Nov., 2000, (TR2000-05)

Divakaran, Ajay, Peker, Kadir, and Sun, Huifang, "Descriptor for spatial distribution of motion activity for compressed video, *Proc. IEEE Int'l Conference on Image Processing*, Vancouver, Canada, September 2000.

Lesh, Neal, Marks, Marks, Patrignani,Maurizio, "Interactive Partitioning,",*Graph Drawing, 8th International Symposium*, Colonial Williamsburg, VA, September 20-23, 2000 (TR2000-31)

Anderson, David, Anderson, Emily, Lesh, Neal, J Marks, Joseph, D Ratajczack, Ryall, Kathy "Human-Guided Simple Search," *AAAI-2000*, August 2000 (MERL TR2000-16)

Divakaran, Ajay, Vetro, Anthony, K. Asai and H. Nishikawa, "Video browsing system based on compressed domain feature extraction," *IEEE Trans. Consumer Electronics*, August 2000.

Ogata, M., Pfister, Hanspeter, H. C. Lauer, Y. Dohi, "A Parallel Pipeline Convolution for Perspective Projection in Real-Time Volume Rendering," *The Journal of the Institute of Image Information and Television Engineers*, Vol. 54, No. 9, 2000

Casey, M. and Westner, W., "Separation of Mixed Audio Sources by Independent Subspace Analysis", in *Proceedings of the International Computer Music Conference*, ICMA, Berlin, August, 2000

Yedidia, Jonathan S., "An Idiosyncratic Journey Beyond Mean Field Theory," in *Advanced Mean Field Methods, Theory and Practice*, eds. David Saad and Manfred Opper, pages 21-36, (TR2000-27)

Lin, J., Divakaran, Ajay, Vetro, Anthony , Kung, S.Y., "Efficient representation and comparison of multimedia content using DAG-composition," *Proc. IEEE Int'l Conference and Multimedia and Exposition*, New York, NY, August 2000

Brand, Matthew, "Exploring variational structure by cross-entropy optimization." *Proceedings, International Conference on Machine Learning*, 2000

Brand, Matthew,V. Kettnake, V, "Discovery and segmentation of activities in video." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (22)8. August 2000

Anderson, David, Anderson, Emily, Lesh, Neal, Marks, Joe, David Ratajczak, David, Ryall, Kathy, "Human-Guided Simple Search", *The Seventeenth National Conference on Artificial Intelligence*, July 30-August 3, 2000, Austin, Texas (TR2000-16)

Frisken, Sarah F. ,Perry, Ronald N. ,Rockwood, Alyn P.,and Jones, Thouis R., "Adaptively Sampled Distance Fields: A General Representation of Shape for Computer Graphics", *Proceedings Siggraph '00*, ACM, July 2000, pp. 249-254 (TR2000-15)

Bhalerao, A., Pfister, Hanspeter, Halle, M., and Kikinis, R.,"Fast Re-Rendering of Volume and Surface Graphics by Depth, Color, and Opacity Buffering," *Journal of Medical Image Analysis* , 4 (2000), p. 235-251.

Brand, Matthew, Hertzmann, A., "Style Machines" *Techical Proceedings*, SIGGRAPH 2000.

Jones, Thouis R., Perry, Ronald N., Callahan, Michael, "Shadermaps: A Method For Accelerating Procedural Shading." Presented at *SIGGRAPH 2001 Conference Abstracts and Applications*. (TR2000-25)

Zwicker, Matthias, Pfister, Hanspeter, van Baar, Jeroen, Gross, Markus, "Surfels - Surface Elements as Rendering Primitives", *Siggraph 2000 Conference Proceedings*, July 2000 (TR2000-10)

Anderson, David, Frankel, Jamie, Marks, Joesph, Agarwala, Beardsley, Paul, Hodgins, Leigh, Darren, Ryall, Kathy,

Sullivan, Yedidia, Jonathan, "Tangible Interaction + Graphical Interpretation: A New Approach to Intuitive 3D Modeling," *Siggraph-2000*, July 2000 (TR2000-13)

Pfister, Hanspeter, Zwicker, Mathius, van Baar, Jeron, Gross, Matt, "Surfels – Surface Elements as Rendering Primitives." *Proceedings of SIGGRAPH 2000*, p. 335-342, New Orleans, LS, July 2000. (TR-2000-10)

Jones, Thouis R., Ronald Perry, Ronald N., "Antialiasing with line samples." Rendering Techniques 2000: *11th Eurographics Workshop on Rendering*, pages 197–206, June 2000. (TR2000-21)

Strohecker, Carol, DeMaso, David Ray, Gonzalez-Heydrich, Joseph, Erickson, Julie Dahlmeier, Grimes, Vellisse Pagan, "The Experience Journal: A Computer-Based Intervention for Families Facing Congenital Heart Disease" *Journal of the American Academy of Child and Adolescent Psychiatry* v39 n6 June 2000 (TR2000-12)

Divakaran, Ajay, S. Sekiguchi, K. Asai and Sun, Huifang, "A description scheme for video based on feature extraction in the compressed domain," *Proc. Int'l Conference on Consumer Electronics*, Los Angeles, CA, June 2000

Vetro, Anthony, Sun, Huifang, and Y. Wang, "An MPEG-4 object-based transcoder: architecture and applications," *Proc. Int'l Conference on Consumer Electronics*, Los Angeles, CA, June 2000

Bao, J.,"Optimization of transceiver systems for broadband fixed wireless access systems." *Proc. Int'l Conference on Consumer Electronics*, Los Angeles, CA, June 2000

Anderson, David., Frankel, Jamie., Marks, Joe., Agarwal, Aseem., Beardsley, Ryan., Hodgins, Jessica., Leeigh, Darren., Ryall, Kathy., Sullivan, Ed., Yedidia, Jonathan S., "Tangible Interactions + Graphical Interpretation: A New Approach to Intuitive 3D Modeling", *Proceedings of SIGGraph 2000*, (TR2000-13)

Vetro, Anthony, Sun, Huifang, and Wang, Y., "Object-based transcoding for scalable quality of service," *Proc. Int'l Symposium on Circuits and Systems*, Geneva, Switzerland, May 2000

Yedidia, Jonathan S., Freeman, William T., and Weiss, Yair, "Generalized Belief Propagation," appeared in *Advances in Neural Information Processing Systems 13*, pages 689-695, (TR2000-26)

Moghaddam, Baback, and Yang, Ming-Hsuan, "Gender Classification with Support Vector Machines," in *Proceedings of the 4th IEEE Int'l Conf. On Face and Gesture Recognition*, FG2000, Grenoble, France, March 2000.

Divakaran, Ajay, H. Ito, Sun, Huifang and Poon, Tommy, "Fade-in/out scene change detection in MPEG-1/2/4 compressed domain," *Proc SPIE Conference on Storage and Retrieval for Media Databases*, San Jose, CA, January 2000.

Divakaran, Ajay, Sun, Huifang "A Descriptor for spatial distribution of motion activity" *Proc. SPIE Conference on Storage and Retrieval for Media Databases*, San Jose, CA, January 2000.

1999

Anderson, David, Anderson, Emily, Lesh, Neal, Marks, Joseph K Perlin, D Ratajczack, Ryall, Kathy "Human-Guided Simple Search: Combining Information Visualization and Optimization," *Proceedings of the ACM Workshop on New Paradigms in Information Visualization and Manipulation*, NPIV'99, November 1999.

Anderson, David, Frankel, Jamie, Marks, Joseph, Leigh, Darren, E Sullivan, Yedidia, Jonathan "Building Virtual Worlds with Physical Blocks," *Proceedings of the ACM Symposium on User Interface Software and Technology*, UIST99, November 1999.

Brand, Matthew, "Voice Puppetry" *Technical Proceedings, SIGGRAPH* 1999.

Brand, Matthew, "Structure discovery in conditional probability models via an entropic prior and parameter extinction." *Neural Computation* (11)5. 1999.

Bao, J., Zeng, S., You, Charles, "A Programmable DSP-based approach to the front-end of demodulators for digital TV receivers", *Proc. International Conference on Signal Processing Application and Technology*, November 1999

Lin, I-J., Vetro, Anthony, Divakaran, Ajay, Kung, S-Y., "Shape representation and comparison using Voronoi order skeletons and DAG ordered trees," *Proc. Int'l Workshop on Very Low Bitrate Video Coding*, Kyoto, Japan, October 1999

Vetro, Anthony, Sun, Huifang, Wang, Y., Guleryuz, O.,"Rate-distortion modeling of binary shape using state partitioning," *Proc. Int'l Conference on Image Processing*, Kobe, Japan, October 1999

Duardo, O., DaGraca, P., Hosotani, S. Sugawa, S., Jiang, H., ``A cost effective HDTV decoder IC with integrated graphics engine and display processor," *IEEE Trans. Consumer Electronics*, vol.45, num.3, August 1999

McCormack, Joel, Perry, Ronald, Farkas, Keith, Jouppi, Norman, "Feline: Fast Elliptical Lines for Anisotropic Texture Mapping." *Proceedings of SIGGRAPH 99*, pages 243-250, August 1999. (TR2000-22)

De, P., Bao, J., Poon, Tommy, ``A calculation-efficient algorithm for decision feedback equalizers," *IEEE Trans. Consumer Electronics*, vol.45, num.3, August 1999

De, P., Bao, J.,Poon, Tommy" Subspace based decision feedback equalization", *Proc. IEEE Radio & Wireless Conference*, Denver CO, August 1999

Pfister, Hanspeter, Hardenbergh, J., Knittel, J.,Lauer, Hugh, Seiler, Larry, "The VolumePro Real-Time Ray-Casting System, *Proceedings of SIGGRAPH 99*, p. 251-260, Los Angeles, CA, August 1999.

Ray, H., Pfister, Hanspeter, D. Silver, T. A. Cook, "Ray-Casting Architectures for Volume Visualization," *IEEE Transactions on Visualization and Computer Graphics*, p. 210-223, Vol. 5, No. 3, July/September 1999. (TR99-17)

Lesh, Neal, Allen, James, "Simulation-based inference for plan monitoring", *AAAI 1999*. The Sixteenth National Conference on Artificial Intelligence, July 18-22, 1999, Orlando, Florida (TR99- 06)

Pfister, Hanspeter, "Architectures for Real-Time Volume Rendering," *Journal of Future Generation Computer Systems* (FGCS), p. 1-9, Elsevier Science, Vol. 15, No. 1, February 1999. (TR99-18)

Project Reports

The body and soul of any research lab is the portfolio of projects it pursues. Therefore it is appropriate that the main body of this annual report consists of descriptions of the various projects being done at MERL. For ease of reference, the reports are grouped into ten topic areas.

- Advanced Digital Television
- Artificial Intelligence
- Audio Video Processing
- Collaborative Interaction
- Computer Vision
- Digital Communications
- Graphics
- Net Services
- Networks
- Spoken Language Interfaces

Each topical section begins with a short discussion of the topic area, highlighting MERL's major efforts. It then continues with a number of one-page project reports. These reports describe projects completed in the last twelve months and major milestones in continuing efforts.
The individual project reports begin with a brief summary at the top, followed by a more detailed discussion. The bottom of the report indicates the principal lab at MERL involved with the project and a contact person. Also included is a characterization of the type of project. The purpose of this is to indicate the kind of result that has been obtained.

- Initial Investigation – Work is underway on the project, but no firm results have been obtained yet. The project report is included to give a better understand of a direction in which MERL is heading.
- Research – The results obtained are in the form of papers, patents, and/or research prototypes. They represent valuable knowledge, but significant advanced development work will be required before this knowledge can be applied to products.
- Advanced Development – The results are (or will be) in forms that can be directly used in product development. The exact form of the result depends on what is being produced. For software projects, the results are typically code that can be directly used in products. For semiconductor chip projects, the results are typically in the form of detailed specifications for algorithms to be embedded in silicon.

Advanced Digital Television

Advanced Digital Television provides an exciting new world for the consumer. Not only does it offer improved picture quality, it also provides a means for seamlessly blending many new services into the TV set, expanding the scope of what televisions can do. Advanced Digital Television also poses new challenges in video encoding, transmission, and reception.

Historically, the Advanced Digital Television effort at MERL has focussed on the development of DTV, HDTV, and related new technologies. While the bulk of DTV standardization is complete, it continues to be upgraded. Hence we continue to monitor standardization as well as implement programming guides (PSIP) and closed captioning systems (DTVCC). We have also been designing, evaluating and field-testing MELCO's DTV receiver. Finally, we have been working on systems that ensure backward compatibility between DTV and traditional standard definition TV. We have focussed on down-conversion of HDTV signals so they can be displayed on a standard definition TV set.

Project Descriptions

Digital Television Standardization

Since the FCC decided on the US standard for digital television at the end of 1996, it has been an uphill struggle in implementing the standard to replace the analog television system, which has been in existence since the early 1950s.

The standard of encoding audio and video is known and there had been great improvements in the encoding algorithm, but the difficulty lies in all the different enhancements and new services that digital brings to multimedia.

Participation in the standardization process greatly benefits advanced product development.

Background and Objectives: As equipment manufacturer of both professional (broadcast encoder) and consumer (receiver decoder) products, we need to be well informed of various standardization in progress; voice either support or dissent in certain technologies or practices according to business needs.

Technical Discussion: Ongoing work includes standards development, updates and implementation issues on DTV Closed Caption (DTVCC), Program System Information Protocol (PSIP), Directed Channel Change (DCC), DTV Application Software Environment (DASE), data broadcasting, and RF modulation enhancements. Enhanced DTVCC is a big improvement compared to the traditional captions for the analog TV. Some compares it to the surround sound for the eyes. Due to the difficulty in implementation for the receiver design, and the lack of broadcast equipment support, DTVCC progress has been slow. Test material has been developed to test receivers under development. PSIP has been a standard since 1998, and since then broadcast implementation has been lagging due to lack of education to broadcasters as well as equipment inter-operability. Work continues in the standard PSIP-encoder interface, and DCC implementation challenges. DASE has been under development within Advanced Television Systems Committee (ATSC) for over 4 years. There are total of 8 parts of the standard, 5 of which has been approved by the technical committee. The entire DASE standard is to be an ATSC standard sometime middle of next year. 8-VSB (Vestigial Side Band), the ATSC modulation standard, has been under discussion whether it is an adequate method of transmission. Current work in progress is to enhance its performance without sacrificing compatibility with current receiver implementation.

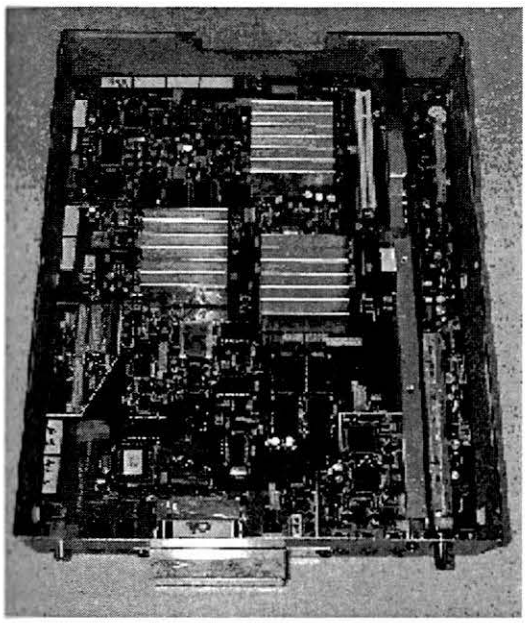
Collaboration: Koriyama Works, Kyoto Works, Eijyoken, Mitsubishi Digital Electronics America, Inc., CBS Network, Consumer Electronics Association.

Future Direction: Koriyama Works, Kyoto Works, Eijyoken, Mitsubishi Digital Electronics America, Inc., CBS Network, Consumer Electronics Association.

Contact: Fernando Matsubara, James Fang
http://www.merl.com/projects/DTVstandard/

Lab: MERL Murray Hill
Project Type: Advanced Development

Digital Television Receiver Module Development



Mitsubishi Digital Electronics America (MDEA) is the foremost leader in consumer digital televisions with over 50% of the market share. To maintain this position, each generation of the new products must incorporate cutting-edge new technology combined with state of the art performance.

Following the history-making first generation product based on the Nozomi Digital TV (DTV) chipset jointly developed by MELCO, MERL-MHL, and Bell Laboratories, the 2002 Mitsubishi DTV will extend its lead in the industry by offering new features not found in any other manufacturers' products.

MERL-MHL, formerly known as Advance Television Laboratory, is contributing to this effort with an unparalleled expertise in DTV signal reception accumulated since the days of its foundation.

Background and Objectives: DTV performance is under constant improvement and the aspect that receives the most scrutiny is the terrestrial digital broadcast signal reception capability. This part of the DTV system for receiving the digital signal is called the "front end," consisting of a tuner and a demodulator. MHL's contribution is in the design, evaluation, and field-testing of the front end culminating to a successful mass production of the DTV receiver module.

Technical Discussion: The tuner and the demodulator each supplied by a different manufacturer need to be integrated and optimized to form a complete front end. Both components are critically chosen and meticulously engineered to function in concert to receive both terrestrial and cable broadcast of digital TV signals. The signal reception performance is evaluated in the laboratory and also through field-testing. While the data gathered is compared against historical as well as competitor's performance levels. Measurements of units from engineering trial production show improvement in all aspects of the front end.

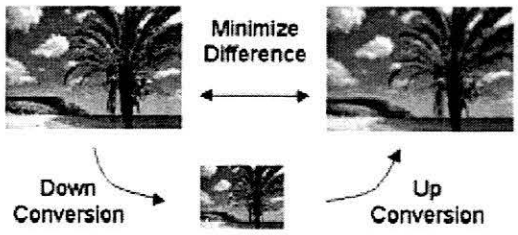
Collaboration: MELCO-Kyoto Works, MELCO-Image Systems Lab, Mitsubishi Microcomputer Software (MMS), Mitsubishi Electric Engineering Company (MEE), Mitsubishi Digital Electronics America (MDEA).

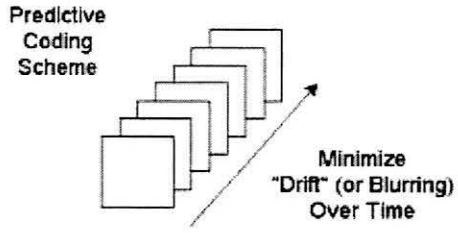
Future Direction: MERL-MHL will continue to contribute to MDEA's effort of developing the current DTV receiver module into mass production in late 2001. Furthermore, MHL is ready to offer its services in planning and development of the next generation of Mitsubishi's DTV.

Contact: George Fang
http://www.merl.com/projects/DM-support/

Lab: MERL Murray Hill
Project Type: Advanced Development

Down-Conversion for Improved Picture Quality





Down-conversion is a fundamental part of the digital television technology that provides compatibility of the emerging HDTV equipment with the traditional standard television broadcasting, and it can also be applied to a variety of other consumer electronics products features or requires variable resolution video image feeds. In this project, we design down/up conversion filters using frequency synthesis and spatial domain methods, and compare the performances of various conversion techniques in terms of image quality. We also investigate the use of extracted side information to improve picture quality and effects on the system architecture.

Background and Objectives: The project is initiated to conduct research on new algorithms for memory reduction within a video decoder. Not only are memory requirements reduced, but implementation costs are also reduced due to impact on memory bandwidth and clock rate. Two key technologies have been the major focus: down-conversion and motion compensation. These new algorithms are basic components of a low-cost video decoder and may be used in a wide range of image/video products.

Technical Discussion: In order to reduce the size an image without causing aliasing errors, the image is filtered initially to limit the bandwidth of the input signal spectrum. A low-pass filter tuned at the Nyquist frequency, which is the half of the signal bandwidth, removes the high frequency coefficients. Then, the filtered image then can be sub-sampled accordingly. The filters used for down-conversion are based on the concept of frequency synthesis, and the filters used to perform the up-conversion are determined by an optimal least-squares solution. The combination of these techniques allows us to achieve significant reductions in the amount of observable drift compared to previously published methods.

Collaboration: Over several years, this project has benefited from collaboration with L-ji-se, Eijo-ken and Johosoken. From 1998-2000, L-ji-se led the development of a 2nd generation HDTV receiver LSI that made use of the down-conversion algorithm, with team members from MERL-MHL, Lucent and Cadence.

Future Direction: We are investigating novel methods to extract and encode side information to improve the quality of edges and high frequency regions in the down-converted sequence with the minimal impact on the current design architecture. We are also looking for new applications, such as transcoding, that can make use of this fundamental technology.

Contact: Anthony Vetro, Fatih Porikli
http://www.merl.com/projects/down-conversion/

Lab: MERL Murray Hill
Project Type: Advanced Development

Artificial Intelligence

As a grand conception, Artificial Intelligence is an area that has been pursued since the earliest days of computer science. Artificial Intelligence focuses on solving problems on which computers perform poorly while people solve almost effortlessly. Examples include common sense reasoning, planning, and learning. Early efforts in Artificial Intelligence focused on understanding human problem-solving by taking on small tasks. The emphasis has now shifted toward building useful systems that exhibit new functionality. These efforts have produced useful sub-technologies, such as expert systems and improved control systems based on computer learning.

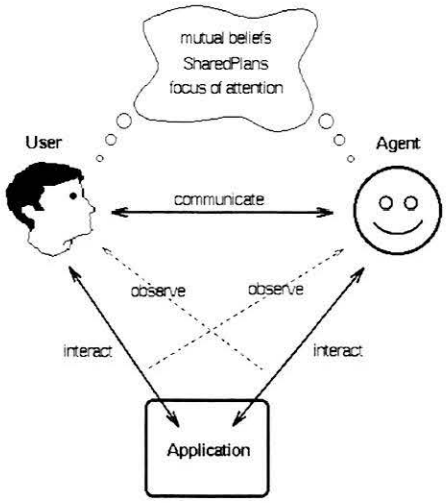
The experience of the last few decades has shown that Artificial Intelligence is often most effective when used as an "ingredient" in other systems. In this way, MERL is using Artificial Intelligence as an ingredient for applications as diverse as user interfaces, digital communications, and computer graphics.

Computer Vision and Spoken Language Interfaces, two technology area commonly considered to be branches of Artificial Intelligence, are listed separately in this MERL Annual Report (see pages 71 and 121).

A significant effort at MERL that builds and expands on technologies developed in the field of Artificial Intelligence is COLLAGEN (for COLLaborative AGENt), object-oriented Java middleware for building collaborative interface agents, described in detail on page 46. Collaborative agents take an active role in helping users solve problems, including correcting errors, suggesting what to do next, and taking care of low-level details.

Project Descriptions

COLLAGEN: Java Middleware for Collaborative Interface Agents



COLLAGEN (for COLLaborative AGENt) is object-oriented Java middleware for building collaborative interface agents. Collaborative agents take an active role in helping users solve problems, including correcting errors, suggesting what to do next, and taking care of low-level details. A collaborative agent can be added to an existing graphical user interface, as illustrated on the left, so that the user and the software agent both communicate with each other and observe each other's actions. COLLAGEN is currently being used in prototype systems for a range of applications.

Background and Objectives: COLLAGEN is based on the study of naturally occurring human collaboration, such as two people assembling a complex mechanical device or two computer users working on a spreadsheet together. We have developed COLLAGEN over the past several years, and are now using it to support a number of research and prototyping activities ranging from operator training and task guidance to collaborative spoken-language interfaces.

Technical Discussion: The heart of COLLAGEN is its representation of the current state of a collaborative conversation, consisting of a plan tree, which tracks the status of the shared application tasks, and a focus stack, which tracks the current focus of attention. COLLAGEN automatically updates these data structures whenever either the user or the collaborative agent produces an utterance or performs an action. The agent uses these structures to determine the appropriate collaborative behavior when it is the agent's turn to speak or act.

Most of the code in COLLAGEN is generic. To apply COLLAGEN to a particular domain, such as VCR programming or gas turbine engine operation, the key information a developer must provide is the task model, which is an abstract formalization of the hierarchical structure of the domain's typical tasks and subtasks.

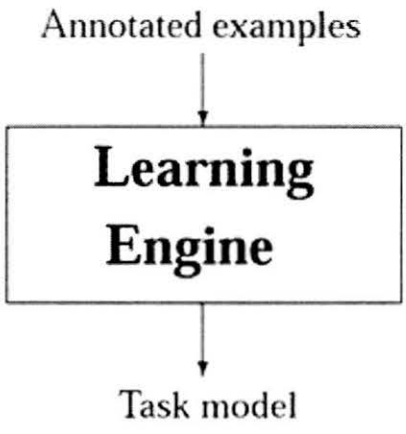
Collaboration: We are currently collaborating with MELCO IESL (Sanken) on the application of COLLAGEN to industrial and automotive systems, with the University of Southern California, Information Sciences Institute and the MITRE Corporation on embedded training applications, with Delft U. of Technology on applications to intelligent consumer products, and with the MIT Media Laboratory on embodied conversational agents.

Future Direction: In addition to continuing to seek new applications (especially involving speech), we plan improvements in the basic operation of COLLAGEN in the areas of turn taking, causal knowledge, and negotiation.

Contact: Charles Rich, Neal Lesh, Candace Sidner
http://www.merl.com/projects/collagen/

Lab: MERL Cambridge Research
Project Type: Research

Learning Hierarchical Task Models by Demonstration



We have developed machine learning techniques for inferring a hierarchical task model from partially-annotated examples of task-solving behavior. Hierarchical task models are used in many artificial intelligence applications, such as intelligent tutoring, plan recognition, planning, and decision theory. Our approach is based on the intuition that it easier for people to demonstrate and discuss concrete examples of how to accomplish tasks than to formalize task model abstractions directly.

Background and Objectives: This work is motivated in part by the COLLAGEN (COLLaborative AGENt) system, which uses hierarchical task models to support a wide range of collaborative human-computer interaction. COLLAGEN embodies a general theory of collaboration, but requires the development of a hierarchical task model for each domain in which it is used. Developing these task models, a special case of the well-known "knowledge acquisition bottleneck," is a significant engineering burden that we hope to lessen by employing machine learning techniques.

Technical Discussion: In our approach, a domain expert first demonstrates one or more sequences of primitive actions required to achieve typical tasks in the application domain. (The primitive actions may be performed directly on an instrumented graphical interface, if such an interface already exists, or by specifying simulated actions for an as-yet-unimplemented application). The expert then reviews and annotates the recorded example sequences. Finally, our learning algorithm constructs a hierarchical task model that generalizes the given examples.

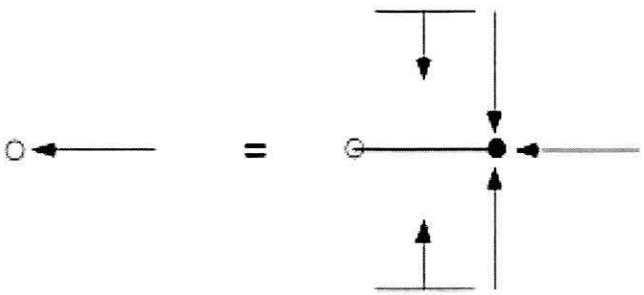
Most previous work on learning task models has dealt only with non-hierarchical models. Our hierarchical learning algorithm also extends previous work by inferring propagators, which enforce relationships, such as equalities, between actions at different levels in the task hierarchy. We have also identified a fundamental search problem, called alignment, faced by any task-model learning algorithm, and have developed both restrictive and preference biases in order to make the alignment search tractable. Our learning algorithm has been proved both sound and complete.

Future Direction: We are planning to create an integrated development environment for task models, which will include our learning algorithms along with facilities for visualizing, editing, and testing hierarchical task models.

Contact: Andrew Garland, Neal Lesh
http://www.merl.com/projects/learning-task-models/

Lab: MERL Cambridge Research
Project Type: Research

Generalized Belief Propagation Algorithms



Many problems in computer vision, machine learning and inference, diagnosis, statistical physics, error-correcting coding, and combinatorial optimization can be posed in terms of a probabilistic graphical model consisting of a lattice of nodes with links connecting nodes that influence each other. Typically, one asks for the probability that a given node or collection of nodes is in some state, given the states of another set of nodes.

Background and Objectives: These problems can be solved exactly when the lattice has the topology of a chain or tree, but in most problems of practical interest, the lattice has loops. We have invented a class of message-passing algorithms which can quickly and accurately (albeit approximately) solve these problems for lattices of arbitrary topology. The algorithms have the attractive feature that one can always pay for greatly increased accuracy with more computation, and the computational cost is not excessive.

Technical Discussion: We discovered that previously developed "belief propagation" (BP) algorithms gave equivalent results to an approximation known in the statistical physics literature as the "Bethe approximation." A class of physics approximations invented by Kikuchi which generalized and improved the Bethe approximation had previously been solved for homogeneous systems like the ferromagnet, but not for inhomogeneous systems such as arise in computer vision. We invented generalized belief propagation (GBP) algorithms that gave results equivalent to the Kikuchi approximations, but which had a form amenable to solution even for inhomogeneous systems. Intuitively, the new algorithms give more accurate results because they allow clusters of nodes to send messages to other clusters, while always maintaining the appropriate constraints between the probabilities in all clusters of nodes. By choosing larger clusters, one can obtain more accurate results, at the cost of more computation.

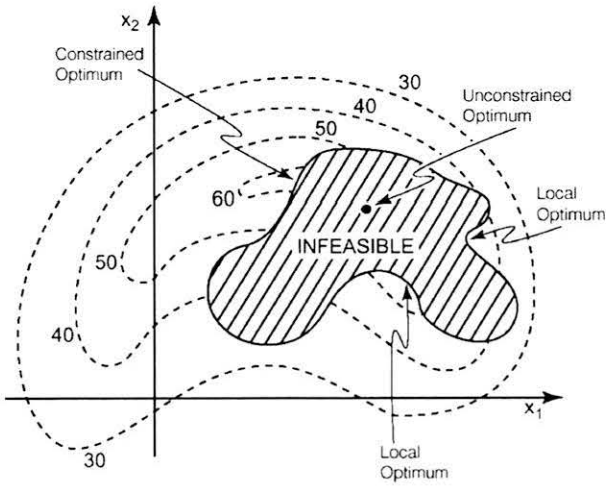
Collaboration: With Yair Weiss (U.C. Berkeley Computer Science)

Future Direction: We have already demonstrated that GBP decoding of error-correcting codes out-performs BP decoding for codes whose graphical representation has short loops, but such codes are not as intrinsically as good as turbocodes and irregular Gallager codes. We are searching for intrinsically very good codes that have short loops. We also hope to apply these new algorithms to other "inference" problems. Essentially any problem that is currently treated by Monte Carlo simulation or simulated annealing might potentially be solved more quickly and accurately by these new algorithms.

Contact: Jonathan Yedidia, William Freeman
http://www.merl.com/projects/GBPalgorithms/

Lab: MERL Cambridge Research
Project Type: Research

A New Method For Numerical Constrained Optimization



Numerical constrained optimization is an important tool with extensive applications in many diverse fields. An ideal problem for constrained optimization is one that has a single measure defining the quality of a solution (the objective function) plus some requirements upon that solution that must not be violated (the constraints). A constrained optimization solver maximizes (or minimizes) the objective function while satisfying the constraints. We have developed a new method for constraint handling that can be applied to established optimization algorithms and which significantly improves their ability to traverse through constrained space.

Background and Objectives: The applicability of optimization methods is widespread, reaching into almost every activity in which numerical information is processed (e.g., science, engineering, mathematics, economics, and commerce). We have focused on solving difficult constrained optimization problems with characteristics common in many engineering applications such as process control (e.g., nonlinear, nondifferentiable, and noisy objective functions and constraints).

Technical Discussion: Many constrained problems have optima that lie near constraint boundaries. Consequently, avoidance of constraints can hinder an algorithm's path to the answer. By allowing (and even encouraging) an optimization algorithm to move its vertices into constrained space, a more efficient and robust algorithm emerges. In the new method, constraints are partitioned into multiple levels. A constrained performance, independent of the objective function, is defined for each level. A set of rules, based on these partitioned performances, specify the ordering and movement of vertices as they straddle constraint boundaries; these rules (employing the insight stated above) have been shown to significantly aid motion along constraints toward an optimum. Note that the new approach uses no penalty function and thus does not warp the performance surface, thereby avoiding the possible ill-conditioning of the objective function typical in penalty methods.

Collaboration: MELCO is involved in many applications requiring optimization, thus providing a potential target for this research.

Future Direction: To continue fundamental research and to embody this research in a product-worthy C library.

Contact: Ron Perry
http://www.merl.com/projects/Optimization-Spider/

Lab: MERL Cambridge Research
Project Type: Research

Audio Video Processing

Most Audio Video Processing technology revolves around industry standards. The Motion Picture Encoding Group (MPEG) is most famous for its digital video compression standards MPEG-1 and MPEG-2. MPEG-1 targets CD-ROM based applications while MPEG-2 targets broadcast quality video. MPEG continues to develop new standards that address the requirements of emerging systems. MPEG-4 is a video compression standard that addresses object-based processing, blending of synthetic and natural video, and low-bit rate transmission, among other things. MPEG-7 addresses standardization of multimedia content descriptions so as to enable applications such as content-based remote and local browsing.

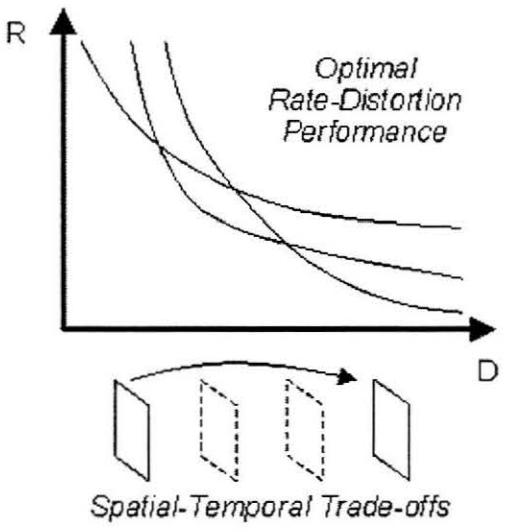
The Audio Video Processing effort at MERL consists of MPEG-4, MPEG-7, and related activities. We have successfully proposed our technologies for adoption into these standards. Our object-based rate control algorithm is now part of the informative (non-normative) part of the MPEG-4 standard. Our Video Motion Activity Descriptor, Directed Acyclic Graph based Description Scheme, Video Transcoding Description Scheme, and Audio Content Indexing and Extraction Method are all now part of the normative part of the MPEG-7 draft international standard. Our current areas of interest include:

- MPEG-4: MERL is developing MPEG-4 Codec technology for encoder optimization and video segmentation. We also have a system development effort for wireless video streaming using the MPEG-4 standard.

- MPEG-7: Our research efforts include technology for video browsing, indexing, and summarization, and surveillance systems, as well as creating new MPEG-7 enabled product and service concepts. Also, MERL is developing technology for the recognition and extraction of individual sound sources from mixed audio scenes.

Project Descriptions

MPEG-4 Video Encoder Optimization



The MPEG-4 standard has been adopted as the solution for mobile multimedia communications and Internet streaming. To improve the quality of video transmission, we consider techniques that optimize the coding efficiency.

We have developed models to estimate rate-distortion characteristics of the video coding process. These models are used to perform bit-rate allocation, maintain a stable buffer, and consider trade-offs between spatial and temporal quality.

Additionally, we focus on developing techniques that reduce the complexity of the video encoding process and ensure robust delivery to client devices over time-varying heterogeneous networks.

Background and Objectives: The topic of encoder optimization has been studied extensively within the context of previous coding standards, such as MPEG-2. In MPEG-4, new coding methods allow for much greater flexibility and the possibility to optimize compression efficiency in different ways.

Technical Discussion: Rate-Distortion (R-D) models are extensively used to solve various encoder optimization problems. We have developed novel R-D models that provide a relationship between the quantizer and rate for coded frames, as well as models that estimate the distortion for coded and non-coded frames. These models have been used by our rate control algorithm to ensure that the encoded video bitstream satisfies various constraints, while maximizing the quality of the reconstructed video. MPEG-4 is the first object-based coding standard and we have developed algorithms to perform bit allocation at the object level. Also, MPEG-4 provides support for variable frameskip; an algorithm that considers optimal spatio-temporal trade-offs has been developed, i.e., code more frames with less quality or fewer frames with higher quality. Finally, new scalable coding techniques in MPEG-4 require advanced methods of modeling, bit allocation and transport; such techniques are currently under investigation.

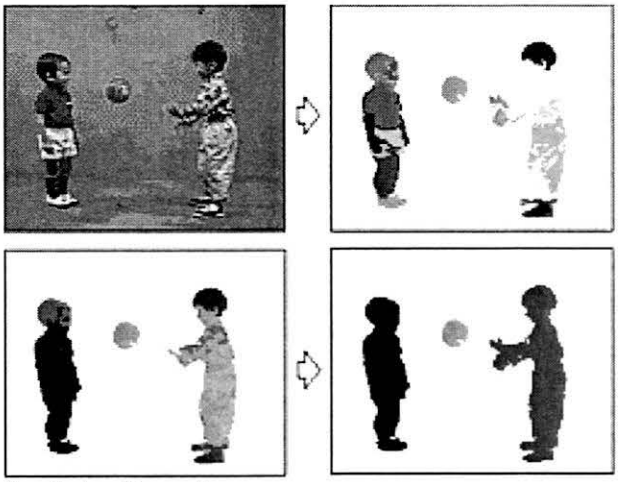
Collaboration: Collaborators include Polytechnic University, the New Jersey Institute of Technology and the Multimedia Information Coding & Transmission Technology Department at Johosoken.

Future Direction: Continue research on improved compression and transport capabilities, with particular focus on scalable coding techniques, R-D modeling and bit allocation.

Contact: Anthony Vetro
http://www.merl.com/projects/MPEG4-encoder/

Lab: MERL Murray Hill
Project Type: Advanced Development

Video Object Segmentation



See Color Figure 1

The increasing demand for multimedia transmission over the Internet and wireless channels requires higher coding efficiency that can be achieved by decoding the shape information of objects. Various multimedia applications such as interactive video annotation, content analysis, summarization, surveillance, video conferencing, and object recognition need extraction of video objects to enable content-based functionalities. Thus, we developed a novel algorithm for automatic and reliable segmentation of moving objects in color video sequences and extraction of video object planes as a part of emerging MPEG-4/7 standards. A novel set of object descriptors are proposed to establish the relation between the different video objects hierarchically.

Background and Objectives: Our method has several advantages over the conventional techniques; it is automatic, computationally efficient, extracts object shape precisely, generates a multi-resolution object tree to expedite content analysis, and able to incorporate priori information.

Technical Discussion: After filtering the input video, markers are selected. Markers serve as the seeds of volumes. A volume is defined as the aggregation of video object planes of the same object in every frame of the sequence. Using the local color and texture characteristics, a volume is grown around each marker. The grown volumes are refined and motion trajectories are extracted. Self-descriptors for each volume, mutual-descriptors for a pair of volumes are computed from trajectories. These descriptors designed to capture motion, shape as well as spatial information of volumes. In the clustering stage, volumes are merged into objects by evaluating their descriptors. Iterative clustering is carried out until the motion similarity of merged objects becomes small. After clustering, an object tree that gives the video object planes for every possible number of objects is obtained. Then, for each segmented region in a frame, a parameterized motion model is fitted from the feature points.

Collaboration: Currently collaborating on MPEG-4/7 video communication technology with Joho-Soken inside of MELCO, and with Polytechnic University, Brooklyn, NY outside of MELCO.

Future Direction: Incorporate MHL video segmentation technology in MELCO CE device, extend video object segmentation applications over the Internet and wireless channels, integrate object detection and tracking into MELCO video surveillance products.

Contact: Fatih Porikli
http://www.merl.com/projects/video-segmentation/

Lab: MERL Murray Hill
Project Type: Research

Robust Optimization-Based Image Reconstruction

See Color Figure 2

There are various applications utilize image reconstruction. In object-based compression techniques, which is becoming more and more crucial for the Internet and wireless video transmission, it can be applied to reduce the textural complexity within the object regions without changing the objects. Reconstruction is especially important for motion and depth estimation problems where it serves as a smoothness constraint that can be easily adapted to the local properties. Decreasing the dynamic color range and requantization are other potential applications. Unlike most filtering techniques (shown here on the right), the robust optimization-based image reconstruction method (shown here on left) refrains from over-smoothing edges and region boundaries. Robust optimization is immune to the image noise, and it does not require the noise to be Gaussian distributed as in the least squares minimization methods.

Background and Objectives: Developing a stable video object segmentation algorithm leaded us to design a noise resistant image filtering scheme. We invented the robust optimization based reconstruction method to prevent from perturbations caused by the common noise filtering algorithms, i.e., low-pass filtering, morphological operators, median filtering, etc.

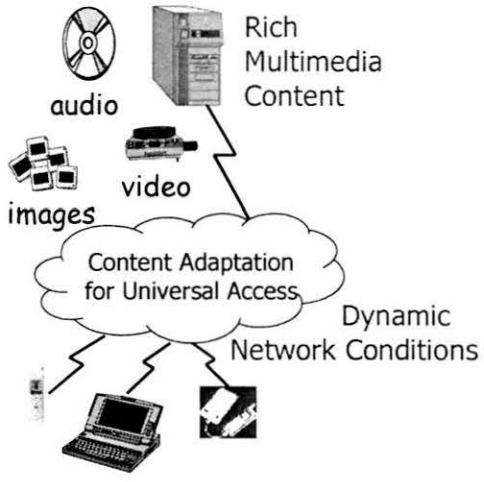
Technical Discussion: In general, the term "robust" is referring to a statistical estimator. It means insensitive to small departures from the idealized assumptions for which the estimator is optimized, in our case the image intensity in a local neighborhood. The word "small" can have two different interpretations, both important: either fractionally small departures for all data points, or else fractionally large departures for a small number of data points. Out of various sorts of robust statistical estimators, we prefer to employ M-estimates that follow from maximum-likelihood arguments. M-estimates are usually the most relevant class for model fitting, that is, estimation of parameters. Given a set of observations, one often wants to condense and summarize the data by fitting it to a "model" that depends on adjustable parameters. We build optimization-based image reconstruction filter by using downhill simplex method to minimize a logarithmic error function. For every image point, a first order model function is fit to the image color channels in a local window.

Future Direction: We are investigating the possible applications of the reconstruction method in image analysis. Future efforts would focus on developing a scalable image representation scheme that enables very low bit rate coding, and promoting reconstructive filters for object recognition.

Contact: Fatih Porikli
http://www.merl.com/projects/optimization-based-filtering/

Lab: MERL Murray Hill
Project Type: Research

MPEG Transcoding Technology



Diverse Set of Terminal Devices

The purpose of a transcoder is to convert compressed content, such as MPEG (Moving Pictures Experts Group) bitstreams, into a form that satisfies transport over dynamic networks, playback/recording on various devices, and users with different preferences. We have developed technology that can assist in the deployment of a wide range of multimedia services, and also help to make system implementation more cost-effective.

For example, sending video to a mobile device poses many constraints, e.g., display, network and power. Transcoding can be used to adapt the content at the server to meet these constraints. Transcoding may also be used in a DVD recorder for the storage of broadcast content with different target bit-rates or spatial resolutions.

Background and Objectives: Recent advances in signal processing combined with an increase in network capacity are paving the way for users to enjoy services wherever they go and on a host of multimedia capable devices. Such devices include personal computers, televisions, and mobile phones. Each of these terminals may support a variety of different formats. Furthermore, the networks that they are connected to are often characterized by different network bandwidth constraints, and the terminals themselves vary in display capabilities, processing power and memory capacity. Therefore, it is required to represent and deliver the content according to the current network and terminal characteristics, while maintaining specified levels of QoS (Quality of Service).

Technical Discussion: In this project, we consider transcoding technology to convert multimedia content according to user, network and terminal characteristics. We have developed conversion methods that transcode MPEG bitstreams to lower bit-rates and spatial resolutions, as well as between formats, e.g., MPEG-2 to MPEG-4. We have also developed summarization methods to provide highlights of the content, e.g., key-frame or key-clip extraction.

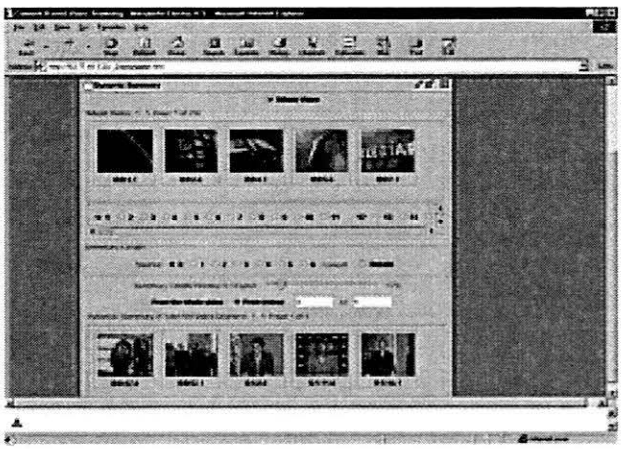
Collaboration: This project is done in collaboration with Princeton University and the Multimedia Information Coding & Transmission Technology Department at Johosoken.

Future Direction: Transcoding is a key technology for many communication systems. In addition to continuing research on transcoding technology, we plan is to collect requirements from business units on how this technology may be used within their systems. It is likely that the technology needs to be adapted to meet certain business needs.

Contact: Anthony Vetro
http://www.merl.com/projects/MPEG-transcoding/

Lab: MERL Murray Hill
Project Type: Advanced Development

MPEG-7 Video Browsing and Summarization



The Murray Hill Laboratory (MHL) of Mitsubishi Electric Research Laboratories has developed a Motion-based Indexing and Summarization system for video. It is based on our work on descriptors of motion activity and their combination with descriptors of other features such as color, as well as our work on Directed Acyclic Graph (DAG) based Description schemes, both of which have also been accepted into the MPEG-7 standard. The purpose of the system is to help the user quickly traverse recorded video content using both top-down i.e. summarization based, and bottom-up, i.e. indexing based, access. We illustrate the MISE web-based interface in the figure on the left.

Background and Objectives: As more and more audio-visual content becomes available in digital form in various places around the world, the ability to locate desired content will become more and more important. Already text based search engines help retrieve textual data from the World Wide Web, but no equivalent identifying information exists for A/V content. The proposed MPEG-7 standard will standardize a multimedia content description interface that will enable efficient searching and browsing of worldwide multimedia content. In this project we emphasize the Personal Video Recorder application, that provides the user with the content he wants when he wants it by storing a large volume of content recorded from broadcast and then providing effective navigation of the stored content using summarization and indexing.

Technical Discussion: The system relies on extraction of compact descriptors in the compressed domain, which makes both the content preparation and the content access fast. It primarily relies on the MPEG-7 motion activity descriptor, and also makes use of simple color histograms. We have a unique motion activity based approach to video summarization.

Collaboration: Joho-Soken.

Future Direction: Our current research focus is on Content Summarization. We plan to incorporate audio feature based indexing and summarization into our framework. We are still working on improving and enhancing the summarization. Our target applications include Personal Video Recorders, Consumer Video Browsing for DVD players and other CE devices, Remote access of video, Surveillance etc.

Contact: Ajay Divakaran
http://www.merl.com/projects/video-browsing/

Lab: MERL Murray Hill
Project Type: Advanced Development

MPEG-7 Contributions

Above: Ten most motion active shots in a news video
Below: Video indexing based on combination of motion and color descriptors

MERL Murray Hill has made three salient technical contributions to the MPEG-7 standard in addition to participating in its creation and shaping. Benefits from such participation in standards include revenue from licensing of proprietary technology included in the standard as well as timely development of expertise related to the standard. We are currently participating in the MPEG-21 standard which provides a framework for exchange of digital media. We are working on specification of user profiles and environments, as well as investigating opportunities for contributing our proprietary technology.

Background and Objectives: As more and more audio-visual content becomes available in digital form in various places around the world, the ability to locate desired content will become more and more important. Already text based search engines help retrieve textual data from the World Wide Web, but no equivalent identifying information exists for A/V content. The proposed MPEG-7 standard will standardize a multimedia content description interface that will enable efficient searching and browsing of worldwide multimedia content.

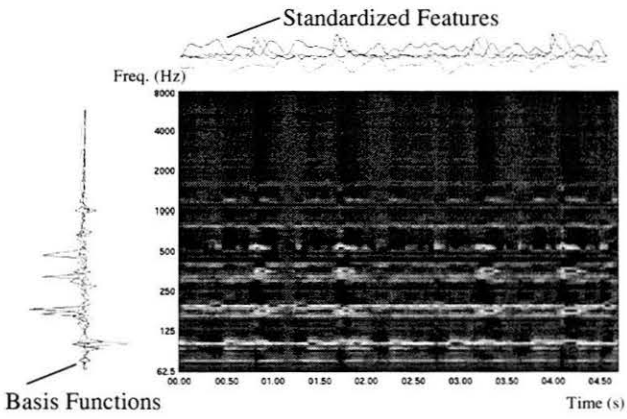
Technical Discussion: Our contributions to the MPEG-7 standard include:
1. The Motion Activity Descriptor - Enables rapid and effective indexing and summarization using the overall pace or intensity of motion in a video segment.
2. Directed Acyclic Graph (DAG) based Description Scheme - Enables content summarization and comparison through multiple interpretations based on various traversals of a DAG-based content description.
3. Transcoding Hints Description Scheme - Meta-data that facilitates simpler and improved quality transcoding of content.

Collaboration: Joho-Soken

Future Direction: Our current research focus is on Content Summarization on one hand and transcoding on the other. We plan to incorporate audio feature based indexing and summarization into our indexing and summarization framework. On the transcoding side we are now working on intelligent use of network resources based on MPEG-7 content descriptions. Our target applications include Personal Video Recorders, Consumer Video Browsing for DVD players and other CE devices, Remote access of video, Surveillance, etc.

Contact: Ajay Divakaran **Lab:** MERL Murray Hill
http://www.merl.com/projects/MPEG7-contributions/ **Project Type:** Advanced Development

MPEG-7 Sound Recognition



See Color Figure 3

As machines are becoming more capable of listening, new tools for identifying sound content are emerging. MERL CRL has developed the sound recognition component of the MPEG-7 international standard for multimedia content description. This enables inter-operable applications such as speaker identification, audio event recognition and automatic music classification. With these tools, audio and video media may be searched much like the text part of a Web page.

The graphic shows a spectrogram of an audio clip and its corresponding MPEG-7 features. Robust automatic sound classifiers may be built using such *signatures*.

Background and Objectives: One of the major challenges in the design of sound recognition systems is selecting features and probability model parameters that are robust across a broad range of sound types. Robust systems should require no human intervention for feature extraction or model parameter estimation. To this end, we sought fully automatic methods for building recognition systems using training data.

Technical Discussion: The MPEG-7 standardized features for sound recognition consist of dimension-reduced spectral vectors obtained using a linear transformation of a spectrogram. Dimension reduction uses the singular value decomposition (SVD) and independent component analysis (ICA) to find a set of basis functions that maximize the information content of the features whilst minimizing their size. Such compact features are essential for efficient training and robust performance.

Within the standard, these features are used in conjunction with hidden Markov models to build robust automatic classifiers. HMM classifiers are represented within MPEG-7 using XML-based description schemes that enable interoperability and portability of models between different applications. The system successfully identifies sound events as diverse as speech, singing, environmental noises, animal sounds, musical instruments and music genres. Industry uses for this technology include remote audio monitoring, media archive searching and automatic music monitoring for broadcast facilities.

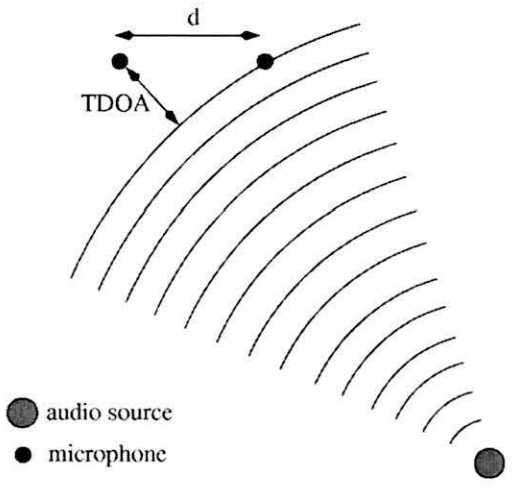
Collaboration: International Standards Organization

Future Direction: We are currently developing a speaker identification system based on the MPEG-7 sound recognition tools. Other planned projects include automatic sound monitoring for surveillance applications and music similarity comparisons for end-user media searches.

Contact: Michael Casey
http://www.merl.com/projects/MPEG7sound/

Lab: MERL Cambridge Research
Project Type: Advanced Development

Robust Sound Localization for Interactive Agents and Surveillance



One crucial requirement for productive research into novel perceptual interface systems is a toolbox of simple, robust perceptual tools. One particularly useful perceptual competence is the ability to localize audio sources in the environment. Progress in the signal processing community has led to the development of a new method for audio localization that is demonstrably superior to the more commonly used Cross-Correlation method. This document introduces this new method, and briefly discusses the inclusion of the algorithm into the prototype embodied perceptual interface agent being developed at MERL.

Background and Objectives: The goal of this work was to create a robust audio localization module to be used in conversational agents and surveillance applications. Special attention was given to performance with respect to speech signals in reverberant environments. A common approach is to calculate the cross-correlation (CC) between the multiple audio. This provides a good estimate of the time difference of arrival (TDOA) for complex signals in the absence of reverberation and micro-periodicities in the source signal. Reverberant environments, such as office spaces will corrupt this estimate of the TDOA, and lead to poor localization performance.

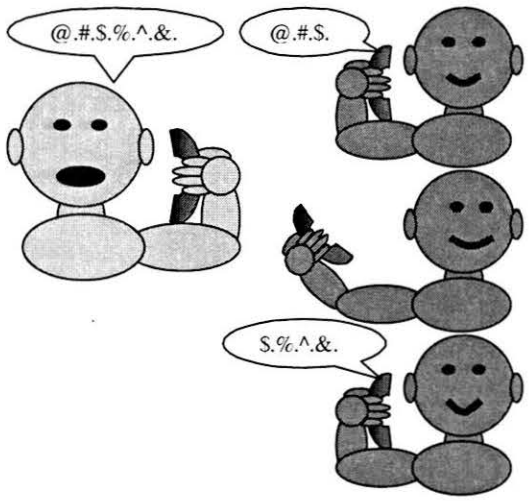
Technical Discussion: An efficient way to compute the CC of two signals is to perform a multiplication in the Fourier domain. This implementation makes it possible to apply a weighting function that de-emphasizes portions of the spectrum that are suspected to be corrupt. This is called the Generalized Cross-Correlation (GCC). One instance of GGC is Cross-power Spectrum Phase, also called the Phase Transform (PHAT). Experiments have shown PHAT to be very robust to noise and reverberation, operating near the theoretical bound on performance of TDOA estimation, as calculated by Gustafsson.

When implemented with the FFTW libraries (fftw.org), the process requires a modest 3% of a 1GHz Pentium III on 16K sample windows at 44KHz sampling rate. A description of experimental procedure is beyond the scope of this document, however we did note marked improvement in audio localization performance after we replaced the CC estimator with the PHAT estimator, especially with respect to reverberation errors.

Future Direction: Integration of audio localization with other perceptual and rational competencies into a toolkit for the construction of embodied conversational agents and next generation surveillance systems.

Contact: Chris Wren, Darren Leigh, William Yerazunis **Lab:** MERL Cambridge Research
http://www.merl.com/projects/robust-audio-localization/ **Project Type:** Advanced Development

Real-Time Audio Buffering for Telephone Applications



On ordinary phones, if you miss something, it is gone forever. With real-time audio buffering, you can hear what you have been missing. Best of all, the operation is intuitive, requiring little or no action on the user's part.

Background and Objectives: The cell phone market is extremely competitive. Many phones have similar capabilities and prices. To stand out in the crowd, unique features are needed. We propose real-time audio buffering as an inexpensive feature users will appreciate.

Real-time audio buffering senses when the phone is removed from the ear. When the phone is returned, the missed incoming audio is played back, but at a rate faster than real-time. Very quickly, the phone catches back up to real-time, and the user misses nothing, despite the distraction. The ear sensor also allows us to create an elegant "instant-replay" function. Briefly move the phone away from the ear, and it automatically replays the last few seconds. Repeating this gesture skips successively further back in time.

Technical Discussion: Real-time audio buffering uses a capacitive proximity sensor to determine when the phone is near the ear. The sense electrode is created by coating the inside of the ear piece with a conductive coating as might normally be applied for shielding purposes. In addition, the remainder of the case interior is similarly coated and connected to the phone's internal ground. Holding the phone up to an ear capacitively couples the ear sense electrode to the phone's ground through the user (via ear and hand). By measuring the change in capacitance, we can easily detect the phone proximity to the ear. The required circuit uses two digital I/O pins and a single capacitor to make the measurement. This is extremely low cost and low power.

The audio section constantly stores incoming audio in a circular buffer. Normally, the "play pointer" immediately tracks the "record pointer". When the phone is removed from the ear, the record pointer continues adding incoming audio, but the play pointer is halted until the phone is returned to the ear. To help give the user the context of the conversation, it helps to move the play pointer a little further back in time each time the phone is removed from the ear. This implements the instant-replay function as a side benefit. When the play pointer is behind the record pointer, we speed up play back via a pitch-preserving technique and skip silent intervals.

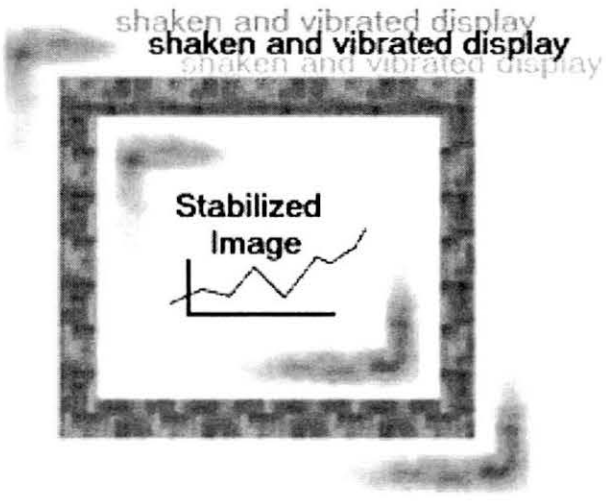
Collaboration: This is a joint effort of MERL Cambridge Systems and Cambridge Research.

Future Direction: We will be doing further user testing to optimize the system.

Contact: Paul Dietz
http://www.merl.com/projects/Audio_Buffer/

Lab: MERL Cambridge Systems
Project Type: Advanced Development

Active Electronic Display Stabilization



It is difficult to use electronic displays on PDAs, cellphones, in-car navigation and entertainment systems, and other personal electronic devices while walking, running, or in a moving or shaking vehicle. This is because the display itself is shaking a distance larger than the size of the characters, so that the eye sees only a blur. We have developed an inexpensive technology to actively compensate for vibration in the display, by measuring the acceleration of the display and displacing the display bitmap an equal and opposite distance, so that the displayed image appears stable even when the physical display hardware is in violent motion.

Background and Objectives: Image-taking stabilization has been widely used to produce better images in telescopes, binoculars, cameras, and camcorders. The use of display stabilization (rather than image-taking stabilization) is new to the commercial market (although there is some prototype military use). We expect the greatest markets for active display stabilization will be for PDAs, cellphones, and in-car navigation and entertainment systems.

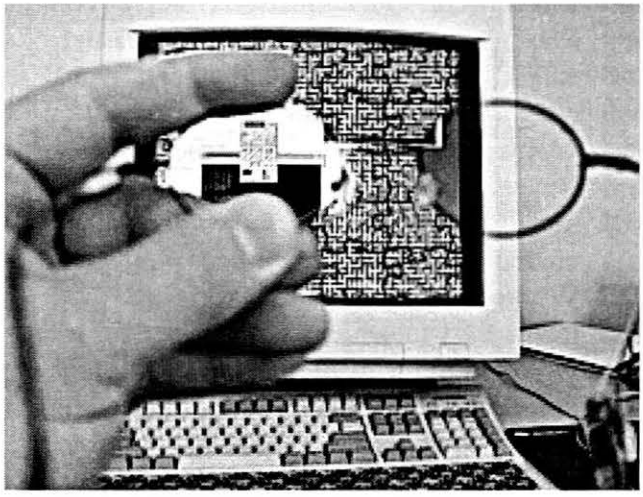
Technical Discussion: The display stabilization system uses an inexpensive micromachined accelerometer to measure the X and Y acceleration of the display surface. The acceleration is digitally integrated to provide velocity information, and high-pass filtered. The velocity information is digitally integrated and high-pass filtered to provide the vibrational displacement information. The desired display bitmap is then offset by the appropriate number of pixels in the -X and -Y directions, so that the image appears unmoving in an inertial frame of reference, even though the physical hardware may be in rapid motion. High-pass filtering is used to prevent offsets and drifts from causing long-term displacements of the display. The entire process is controlled by an inexpensive 16-bit microcontroller (a Mitsubishi M16C). The current apparatus uses a 35 x 35 LED array as a display. We are exploring how to perform similar stabilization on liquid crystal and CRT displays.

Future Direction: Extension to LCD and CRT display technology; integration into display drivers.

Contact: William Yerazunis, Darren Leigh
http://www.merl.com/projects/shaky-display/

Lab: MERL Cambridge Research
Project Type: Initial Investigation

Privacy Enhanced Computer Display



The privacy-enhanced computer display uses a ferroelectric shutter glasses and a special device driver to produce a computer display which can be read only by the desired recipient, and not by an onlooker. The display alternately displays the desired information in one field, then the inverse image of the desired information in the next field, at up to 120 Hz refresh. The ferroelectric shutter glasses allow only the desired information to be viewed, while the inverse image causes unauthorized viewers to perceive only a flickering gray image, caused by the persistence of vision in the human visual system (as illustrated here). It is also possible to use the system to "underlay" a private message on a public display system.

Background and Objectives: The Privacy Enhanced Computer Display is appropriate for situations where a computer display is to be located in a publicly accessible area, but the display itself may show information that should be kept private. Places such as banks (bank balances) hospitals (patient health information), pharmacies (drug information), airline ticketing and airport gate agent stations (passenger and security information) are all candidates for a privacy-enhanced computer display.

Technical Discussion: The system consists of two display sources (the public display image P_{ij} and the secret display image S_{ij}) a special device driver, a fast-refreshing CRT, and a set of synchronized ferroelectric shutter glasses. If there is no desired public display, the RGB value (255, 255, 255) can be used. The special device driver alternately displays images of $(P_{ij} - S_{ij})$ and (S_{ij}). The human eye's persistence of vision blurs the two images into $(P_{ij} - S_{ij} + S_{ij}) / 2$, which reduces to $P_{ij} / 2$, which is the public image, without any secret image being visible. To someone wearing the synchronized glasses, only the S_{ij} image is seen.

Collaboration: We are seeking collaboration for development in computer security systems.

Future Direction: The current system uses wired shutter glasses; we will consider upgrading to infrared at some time in the future. We will also explore using other display technologies such as LCD and DLV (digital light valve) systems.

Contact: William Yerazunis
http://www.merl.com/projects/privatedisplay/

Lab: MERL Cambridge Research
Project Type: Initial Investigation

Collaborative Interaction

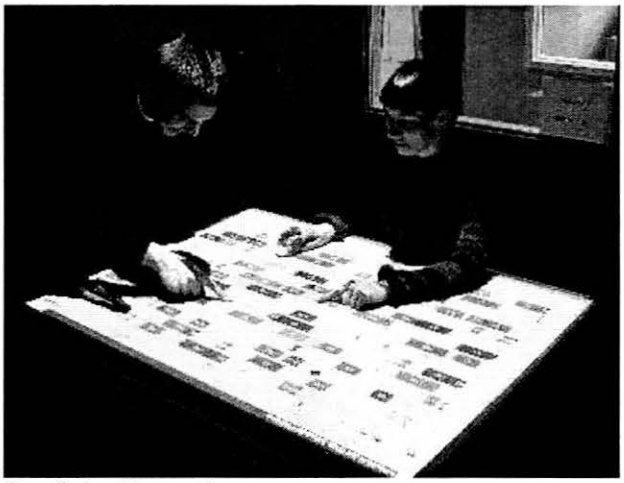
Collaborative Interaction is a technology area that uses advanced computer hardware and software to help people collaborate with each other. MERL is pursuing many different approaches to improving Collaborative Interaction.

We are exploring ways that computation can support face-to-face collaboration and communication between people, either to solve problems or simply to share experiences. Similarly, we are exploring ways in which people can interact with computers in more "natural" ways than are afforded by the traditional interface devices such as keyboards, mice, and monitors. Our goal is to help computers "disappear," as have previous successful technologies, such as writing and electricity, so that people can use computers easily without being distracted by the technology itself.

This is exemplified by our work with projector-based systems that create work spaces for several users, such as the Personal Digital Historian and DiamondTouch. In addition we are developing computer-based collaboration technologies, such as the Collaborative web browser.

Project Descriptions

Human-Guided Tabu Search



See Color Figure 4

We have extended past work on human-in-the-loop optimization, allowing human users to guide a more powerful search algorithm, called "tabu" search.

There are two advantages of tightly integrating human judgement in the process of optimization. First, the human users can guide systems to solutions that satisfy various real-world constraints. Second, human users can use their visual skills, ability to learn, and strategic sense to improve the performance of the computer search algorithm.

Background and Objectives: The motivation of this project is to combine the skills of people and computers to better solve hard combinatorial problems, such as scheduling or vehicle routing problems. In our previous work, users could invoke only a simple, exhaustive algorithm, known as "hill-climbing". We have extended our approach so that users can also guide a more powerful, heuristic search.

Technical Discussion: In our system, the users control the computer's search algorithm by assigning "priorities" to elements of the current solution (e.g., operations in jobshop scheduling, or customers in vehicle routing problems.) These priorities constrain the set of possible modifications that the search algorithm will consider, thus allowing the user to focus the search algorithm on promising areas of the search space. We have developed a version of tabu search that can be constrained by the same priority mechanisms as exhaustive search. Furthermore, this algorithm can itself be encoded as a function which continually adjusts the priorities, thus making tabu search easy to implement in our systems. Our initial results indicate that human-guided tabu search is superior to unguided tabu search. And, users can find significantly better solutions when given tabu search in addition to exhaustive search. Using our new system, we have already found new best solutions to benchmark vehicle-routing problems found on the web.

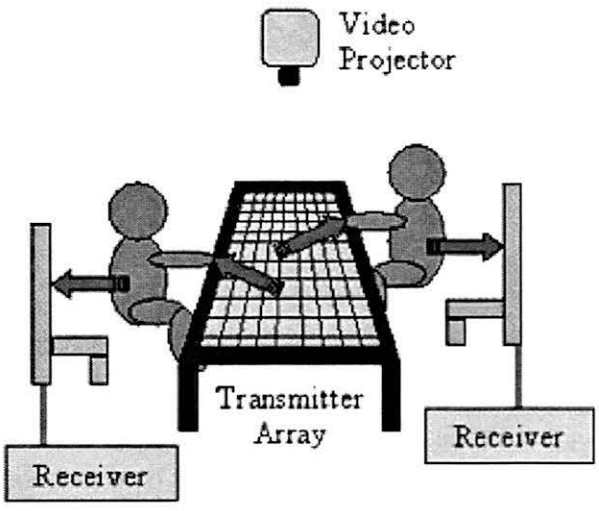
Collaboration: We are adapting our current prototype to work on real-world vehicle routing problems, provided by Sanken.

Future Direction: We are both improving and evaluating human-guided tabu search on multi-depot vehicle routing problems. We are trying to develop ways for a human user to better control and visualize the tabu search, since it is somewhat more complicated than the simple, exhaustive search that we have previously used.

Contact: Neal Lesh
http://www.merl.com/projects/GuidedTabu/

Lab: MERL Cambridge Research
Project Type: Research

DiamondTouch



DiamondTouch is a touch technology which supports multiple, simultaneous users. It is inexpensive, durable, extremely tolerant to debris and offers the unique feature of being able to identify who is touching where.

Background and Objectives: Computer-assisted collaborative environments allow multiple users to work together to solve difficult problems. However, simply scaling the techniques that work well for a single user creates awkward systems. In a collaborative environment, sharing a single mouse is very awkward. Providing multiple mice only makes matters worse - it is extremely difficult to keep track of what each user in the group is doing. Touch screen technologies appeared promising since a glance reveals each users activities. However, touch technologies are usually restricted to detecting a single touch point. They also tend to be somewhat fragile. DiamondTouch is a touch technology specifically designed for a multi-user environment. It detects multiple, simultaneous touches, and even identifies which user is touching where. In addition, it is extremely durable, scales well to large sizes, and is relatively inexpensive.

Technical Discussion: DiamondTouch works via capacitive coupling through the users. The touch surface consists of a series of independent antennas, each transmitting a unique signal. When a user touches near an antenna, the signal from that antenna is capacitively coupled through the user to a receiver attached to the user's chair. Since each chair has a separate receiver, the system can independently determine where each user is touching.

Our prototype system uses a diamond-shaped row-column antenna pattern, which can uniquely determine a single touch point for each user simultaneously. If any user touches multiple points, we determine the bounding box containing all of the points. A full x-y grid of antennas would allow us to uniquely determine all points touched by each user, but this is considerably more expensive and unnecessary for most applications.

Collaboration: This project is a joint effort of MERL Cambridge Systems and MERL Cambridge Research.

Future Direction: We are currently building a large scale prototype which we hope to publicly demonstrate at UIST2001 in November. In addition, we are exploring various extensions of the basic idea, including the feasibility of transparent versions for rear projection applications.

Contact: Paul Dietz
http://www.merl.com/projects/DiamondTouch/

Lab: MERL Cambridge Systems
Project Type: Research

Personal Digital Historian (PDH) for Home and Business

We are developing a novel single-display multi-person system for people to share visual and audio information in face-to-face settings. The goal of the PDH project is to enable multi-modal, interactive and exploratory usage of digital data including photos, audio and video by multiple participants. This technology can be applied to both business meetings (e.g., managers discuss project planning), and social gatherings (family or friends share stories). The project includes research in the areas of content annotation, retrieval, presentation, visualization of and user interaction with images, audio, video and data.

Background and Objectives: As part of people's daily life at work and at home, their computers and digital cameras generate tremendous amounts of digital data. However, technologies that allow people to easily utilize this digital data in a face-to-face conversational or group setting are lagging far behind. Applications are limited by the user interface potentials of current desktop computers and handheld devices. The objective of the PDH project is to take a step beyond.

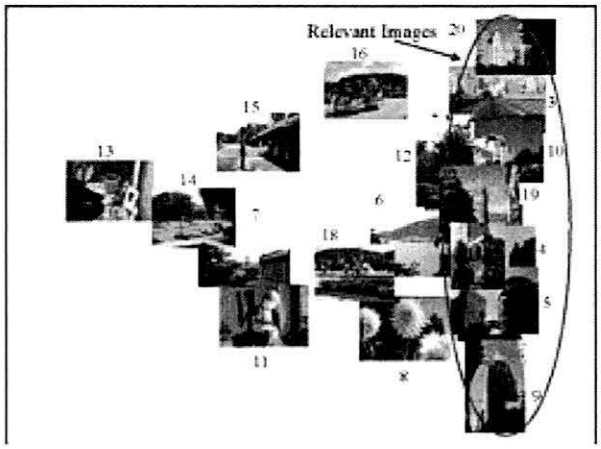
Technical Discussion: The following are the design principles of the PDH system: (a) the user interface will enable natural face-to-face conversation - not forcing everyone to face in the same direction or at their own version of the data; (b) the physical display device must be convenient and customary to use - helping to make the computer disappear; (c) the system should enable interactive and exploratory usage scenarios. The initial PDH prototype employs a circular user interface with top projection onto a physical table surface. Current research focuses on three specific areas: (1) UI and HCI: User interface on non-rectangular real world objects, easy re-orientation of both the user interface and the content, Cartesian to Polar coordinate visual layout translation algorithms, radial deformation to display large amount of images/contents, multiple input and manipulation points, and natural visual and audio query formulation with minimal menu-driven interaction. (2) Content organization and annotation: Use 4 W's (who, where, when, what) from storytelling as the primary annotation for organization and navigation, other semantic annotation include keywords and captions, as well as automatically extracted low level visual feature vectors (color, texture, structure). Contents will be annotated in MPEG-7 format. (3) Content-based image layout: Smart layout, and user-guided image layout, using feature weighting vector estimation in high-dimensional feature space, and then projecting the images from high-dimensional feature space to 2-dimensional displaying space by Principle Component Analysis (PCA), with post linear and/or nonlinear processing.

Future Direction: Develop algorithms for (1) summarizing large sets of data, (2) generating suggestions for information that is related to the users recent queries, and (3) combining semantics, visual features and audio for retrieval and layout. Carry out user studies on annotation of one's own content vs. other people's or professional content. Evaluate and improve the prototype PDH system through user testing.

Contact: Chia Shen, Neal Lesh
http://www.merl.com/projects/PDH/

Lab: MERL Cambridge Research
Project Type: Research

Visualization & Layout for Image Libraries



We have developed a prototype system for visualization and layout with an intuitive browser for retrieval and navigation in large photo libraries. Optimized layouts reflect mutual similarities as displayed on a 2-D screen, hence providing a perceptually intuitive visualization as compared to traditional sequential 1-D content-based image retrieval systems. A framework for user modeling also allows our system to learn and adapt to a user's preferences.

Background and Objectives: Traditional image database retrieval systems display query results as a list, sorted by similarity to the query. This presents one major drawback: relevant images can appear at separate places in the ordered list. The purpose of our proposed content-based visualization is augmenting a user's perception so as to visualize a large information space that cannot be easily perceived by traditional sequential array. The retrieved images are displayed not only in ranked order of similarity from the query but also according to their mutual pair-wise similarities, so that similar images are grouped together (see figure).

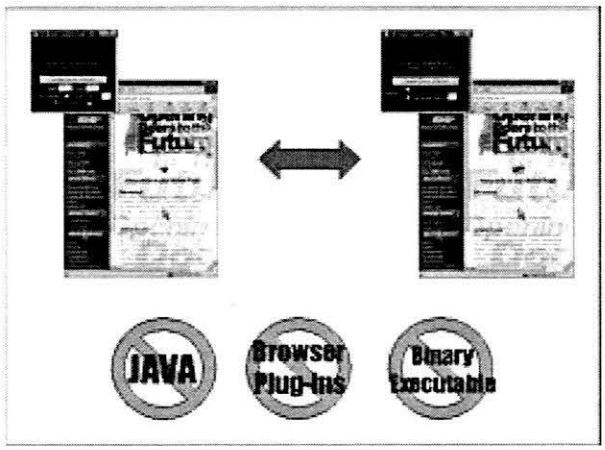
Technical Discussion: Our system represents an image as a 37-dimensional vector of visual features corresponding to color, texture and structure. An ensemble of images can then be projected to the 2-D screen based on Principle Component Analysis (PCA). PCA is a very fast linear transformation that achieves the maximum distance preservation from the original high dimensional feature space to 2-D space among all the linear transformations. This visual layout is denoted as a "PCA Splat" (see above figure). The layout is further optimized for maximal visibility by adjusting the size and position of each image to minimize overlap while relating other dimensions of information (such as relevance/similarity to the query). The mathematical technique used with "PCA Splats" easily lends itself to modeling a user's style and preference. This is achieved by using a weighted subspace projection model which automatically computes the relevance of the key visual features such as color, texture and structure in a given layout.

Future Direction: The system described above is one of the core image retrieval and visualization engines of MERL-CRL's "Personal Digital Historian" (PDH) project. Future areas of research include hybrid visual/semantic models, automatic summarization of large collection of images and user support and automatic aid in performing high-level tasks such as story-telling.

Contact: Baback Moghaddam, Neal Lesh, Chia Shen
http://www.merl.com/projects/vizlayout/

Lab: MERL Cambridge Research
Project Type: Research

Collaborative Web Browsing



The goal of the Collaborative Web Browsing project is to enable simple, real-time, distributed collaboration via web browsers. This effort allows Internet users to collaborate (without any special software) by remotely controlling each other's web browser. The end user requirements are very lightweight to insure that users can participate simply by visiting a web page. Despite the minimal system requirements, a surprisingly high level of collaborative functionality is provided -- such as distributed browsing, scrolling, form-filling, and pointing.

Background and Objectives: Real-time, browser-based collaboration allows users anywhere on the Internet to synchronously view the same content, at the same time, using their web browsers. For example, if one user browses to a new web page, or scrolls down in a web page, then every other user in the collaborative session will see that change reflected in their web browser, too. It is desirable to provide such a collaboration solution that allows arbitrary users to collaborate immediately from anywhere using a standard unmodified web browser. Some heavyweight solutions require users to pre-install a binary application or a browser plug-in. Existing solutions may be proprietary, and may have significant network bandwidth requirements. In order to leverage existing web sites and web tools, it is also desirable to find a solution that allows users to collaborate on web content. That is, users should be able to see and manipulate actual (HTML) web pages in their browsers, rather than being restricted to content that is in some proprietary data format. Furthermore, web page authors or web site administrators should not have to modify the contents of their existing web pages in order to allow users to collaborate on them. Nor should users have to learn a new user interface. Finally, it should optionally be possible to conduct such collaborative sessions privately, and through existing firewalls.

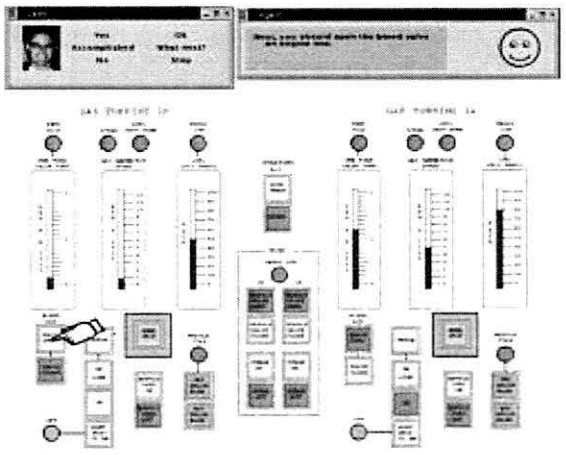
Technical Discussion: Our solution to provide browser-based, real-time, distributed collaboration that meets all of the objectives outlined above involved keeping things extremely simple on the client side. A separate browser window or frame contains a monitor control panel web page with Javascript code that regularly detects changes in the target browser window or frame (in which the target web page being used for collaboration is loaded). Users can initiate collaboration simply by visiting a web page or activating a Bookmark/Favorite to load the monitor control panel. It is assumed that users are talking on the telephone while collaborating, although it is straight-forward to integrate this solution with Internet phone applications.

Future Direction: Future efforts will focus on issues such as robustness, field-testing, deployment, alternative environments, security, multiple sessions, and administrative interfaces.

Contact: Alan Esenther
http://www.merl.com/projects/cwb/

Lab: MERL Cambridge Systems
Project Type: Initial Investigation

Intelligent Agents For Operator Training and Task Guidance



Intelligent software agents can help the operators of complex industrial equipment both with training and in the actual performance of their tasks, by guiding them through the steps of complex procedures, giving positive and negative feedback on their actions, and adapting to their skill level. We have produced generic software for building such software agents and have demonstrated it by building an agent for training operators of a simulated gas turbine engine.

Background and Objectives: This work extends the COLLAGEN (COLLaborative AGENt) system for building collaborative interface agents, treating training and task guidance as points along a spectrum of human-computer collaboration. Our goals are to enhance existing training systems by adding more sophisticated tutorial strategies and to build training and task guidance systems for new applications. Potential application areas include supervisory control and data acquisition (SCADA) systems, such as industrial plant control centers, equipment maintenance tasks, such as elevator repair, and the use of complex software, such as computer-aided design tools.

Technical Discussion: Our approach to operator training is based on "learning by doing."

The software agent guides the operator through a sequence of example scenarios that incrementally expose the operator to the full complexity of the task to be learned. During the training process, the agent maintains a model of the operator's proficiency in each part of the task, so it can appropriately introduce new subtasks as well as give the operator the opportunity to practice previously taught knowledge.

Using COLLAGEN as our implementation base gives the training system developer two major advantages. First, we have an application-independent architecture: to build an agent for a particular application, the developer only needs to provide a hierarchical task model and a goal/initial-condition pair for each example scenario. All of the tutorial strategies are coded into a generic training agent. Second, the same hierarchical task model can be reused to obtain an agent that will collaborate with an operator during the actual performance of the task, if desired.

Collaboration: We are working closely with embedded training researchers at the University of Southern California, Information Sciences Institute and the MITRE Corporation.

Future Direction: We are planning to improve the robustness of our current prototype for gas turbine engine training to support a pilot user study.

Contact: Neal Lesh
http://www.merl.com/projects/training/

Lab: MERL Cambridge Research
Project Type: Research

Computer Vision

Computer Vision combines sensors with data processing algorithms to perform "eye" functions for electronic products and systems. Computer vision can perform functions too tedious or time consuming for humans, such as visual inspection of parts or detection of visual anomalies. Computer Vision is also a key element in developing computer interfaces that are natural and easy-to-use.

We are building technology involving recognition and interpretation of the human form, including poses and motions related to the eyes, face, head, hands, and the entire body. Applications include surveillance/biometrics, user interfaces for games or home appliances, and safety systems for automobiles. Our work in the surveillance area includes Computer Human Observation (CHO) projects (e.g., Fast Face Detection and Stereo Computer Vision) and a proposed Surveillance Architecture.

A further area of research is on the recovery of 3D structure from an image sequence. We are using an integrated system to recover 3D models composed of points, straight line segments, and/or planes. The system is automatic and robust. The goal is to make complex 3D model generation – using still images or video – simple, making the 3D graphics technology more commonplace.

Project Descriptions

Fast super-resolution method

See Color Figure 5

For a variety of applications, we want to be able to increase the resolution of images. The ideal algorithm should be fast, and should add sharpness and detail, both at edges and in regions of texture, without adding artifacts. We have made a new version of our Markov network super-resolution algorithm which does not rely on the iterative belief propagation algorithm. The new algorithm has several speed improvements which result in doubling the resolution of an image from 100x100 to 200x200 in less than 2 seconds.

Background and Objectives: For display of images on high resolution display devices, it is desirable to have an algorithm that increases the resolution of the displayed image, so that it has a more pleasing appearance. We would like to achieve higher image quality than our competitors by using a high quality, machine-learning-based image resolution enhancement algorithm.

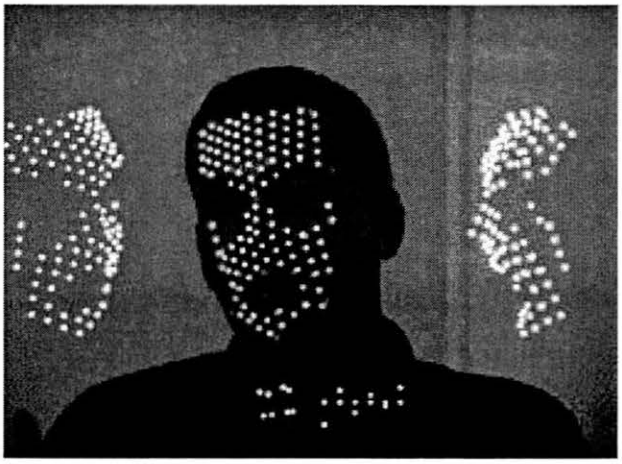
Technical Discussion: We use a training based approach. We examine many pairs of high resolution, and low resolution versions of the same image data. We divide each image into patches, both high resolution and low-resolution patches. We form a training database of 50,000 - 100,000 high and low-resolution patches.<p>Given a new low-resolution image, we seek to estimate the most probable corresponding high-resolution image. In the training database, there may be several different examples similar to any given input low-resolution patch. However, when we include the boundary conditions corresponding to the neighboring patches that have already been selected by the algorithm, we can find a single best-fitting high-resolution patch for each position in the resolution-enhanced image. We structure the database as a tree, and keep a list of the nearby neighbors to each leaf of the tree. This allows for a fast approximation to the nearest neighbor search that we seek to perform for each patch with the entire dataset. This efficiency results in the dramatic speed-up over our previous implementation of the algorithm.

Future Direction: We hope to decrease still the time required for processing the images, and to explore the effect of the training set on the image quality of the resolution enhanced image.

Contact: William Freeman, Thouis "Ray" Jones
http://www.merl.com/projects/fastsuperres/

Lab: MERL Cambridge Research
Project Type: Research

Visual tracking of flexible 3D surfaces



The flexible tracker is a single-camera computer vision system that estimates the 3D translation, rotation, and deformations of nonrigid surfaces such as faces. The input is low-resolution low-quality video and the output is a just a handful of parameters encoding the motion and deformation. This is ideal for animation and for highly compressed video transmission.

Background and Objectives: Reliable tracking of faces and other nonrigid 3D objects will be an important enabling technology for entertainment (animation and special effects), communication (low bit-rate video coding and related MPEG-4) and surveillance (people tracking, biometrics, and resolution enhancement). Our goal is to produce a tracker that works with a single low-resolution low-quality camera and yet produces high-quality 3D tracking information, including precise coding of surface deformations such as the motion of an eyebrow.

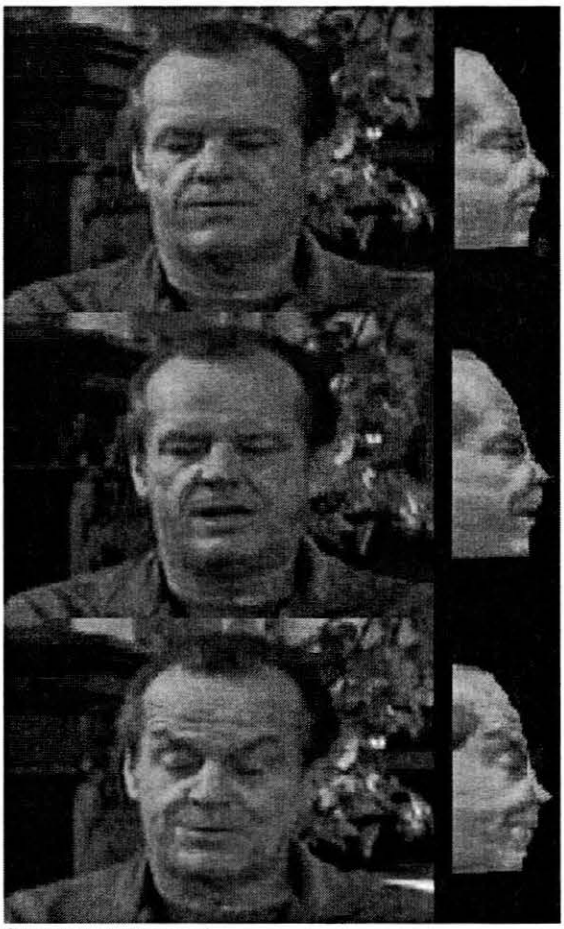
Technical Discussion: We developed a linear framework for model-based tracking of nonrigid 3D objects and for acquiring such models from video. 3D motions and deformations are calculated directly from image intensities without information-lossy intermediate results. Measurement uncertainty is quantified and fully propagated through the inverse model to yield posterior mean and/or mode pose estimates. A Bayesian framework manages uncertainty, accommodates priors, and gives confidence measures. We obtained highly accurate and robust closed-form motion estimators by minimizing information loss from non-reversible (inner-product and least-squares) operations, and, when unavoidable, performing such operations with the appropriate error norm. For model acquisition, we can refine a crude or generic model to fit the video subject. Demonstrated uses are 3D tracking, 3D model refinement, and super-resolution texture lifting from low-quality low-resolution video.

Future Direction: We are now developing methods to handle camera distortion, high-speed tracking tasks, automatic registration of the 3D model to the video, and tracking-based control of an animated "digital actor" (also known as "virtual idol").

Contact: Matthew Brand
http://www.merl.com/projects/FlexFlow/

Lab: MERL Cambridge Research
Project Type: Research

3D morphable models from video



See Color Figure 6

This technology extracts 3D models of nonrigid surfaces (such as faces) directly from intensity variations in video. The resulting model is morphable, meaning that it contains information about the natural ways in which the object changes shape, for example, how the facial tissues move as one speaks. The model can be used for animation, special effects, biometrics, security, and CAD. The image at left shows 3 of 150 video input frames and corresponding profile views that are synthesized from the recovered 3D information.

Background and Objectives: Automatic acquisition of 3D shape and texture is an intensely studied problem due to the growing demand for 3D models in industries ranging from entertainment to engineering. Current methods assume that the object is rigid and require specially calibrated equipment with multiple cameras. Our results are the first general solution for nonrigid objects and ordinary uncalibrated video.

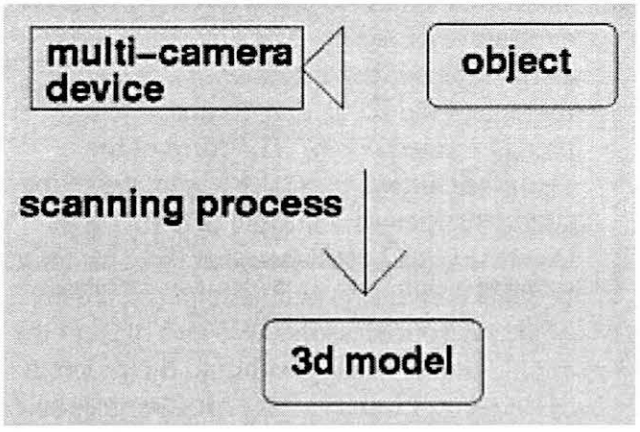
Technical Discussion: We have developed a linear framework for recovering the 3D shape, motion, and articulations of a nonrigid 3D object from a single stream of video. The output is a 3D morphable model and 3D motion/morph parameters, suitable for animation or video rendering. These methods are intended to work with low-quality low-resolution sources such as low-end consumer video, so there is a formal emphasis on robustness. There are two main results: First, we developed a 2D method for tracking surfaces in video that ensures that the motion is consistent with a low-dimensional space of nonrigid motions. Second, we developed a solution for factoring the image intensity variations associated with these 2D motions into 3D shape, rotations, and deformations.

Future Direction: We are currently looking at extending these methods to handle distortion due to camera lenses and strong perspective effects.

Contact: Matthew Brand
http://www.merl.com/projects/3Dvideo/

Lab: MERL Cambridge Research
Project Type: Research

A Hand-Held Multi-Camera Device for 3D Scanning



Creating 3D models from images is an increasingly common technology. Current systems rely on active methods such as light-striping, or make use of a turntable with fixed cameras, or require calibration patterns in the background, or require manual input. This research is on a multi-camera device which is hand-held, which can be moved around an object of interest in an arbitrary way, and which is fully automatic in creating a 3D model. The initial prototype does utilize a calibration pattern close to the object of interest, but the approach evolves naturally into a system which does not require a calibration pattern. That will be the subject of the next phase of the work.

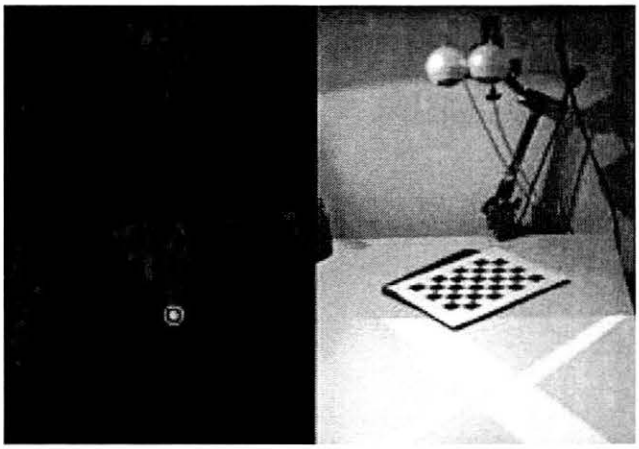
Background and Objectives: The existing market for 3D scanners is increasing in size. Current systems can be expensive (active systems such as laser-based systems), or bulky (turntable systems), or time-consuming (systems requiring manual input). This research is aimed at creating a compact hand-held 3D scanner, which is easy-to-use, and is cheap enough for the consumer market.

Technical Discussion: A key component in the architecture of any passive 3D scanning system is the determination of camera position relative to the object of interest. Subsequent stages then involve computation of surface points on the object, sometimes via identification of the silhouette, and generation of a 3D representation of the surface. The main obstacle to having a hand-held scanner, and the subject of this research, is the first component - automatic and robust recovery of camera position for a scanning device which is being moved arbitrarily around an object of interest.

Contact: Paul Beardsley
http://www.merl.com/projects/3dscan/

Lab: MERL Cambridge Research
Project Type: Research

A Fast Algorithm for Depth Segmentation



This document presents a fast algorithm for depth segmentation. The algorithm uses pre-computed disparity maps to detect regions of the scene that do not have a predetermined depth. The form of the algorithm allows it to take advantage of the single instruction, multiple data (SIMD) instruction set extensions that have recently become commonly available in consumer-grade microprocessors. We also present the application of the algorithm to the problem of detecting contact between a foreground object (a hand) and a geometrically static background (a display surface) in real-time video sequences. Due to the geometric nature of the segmentation algorithm, the touch detector is invariant to lighting, color, and motion. It therefore is applicable to interactive projected displays. A simple extension to the algorithm allows for the background surface to be given analytically.

Background and Objectives: The increasing availability of cheap cameras and high-quality projectors is enabling a wide array of novel interactive space applications. These interfaces present a particular challenge for computer vision algorithms since surfaces lit by bright, high-contrast projection displays may have a continuously changing visual appearance. This means that standard background subtraction techniques that relay on static background appearance will not work. Even worse, front projected displays also cast light on foreground objects, making color tracking and other appearance-based methods difficult or impossible to use.

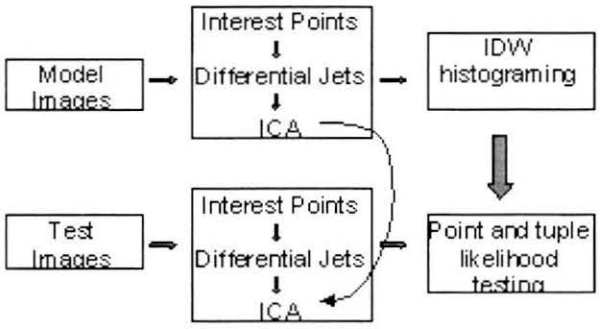
Technical Discussion: One constraint that remains in these situations is the geometric configuration of the projection surfaces. A feature that exposes this constraint is stereo disparity. It is possible to utilize disparity for segmentation without computing a dense depth map by instead employing pre-computed disparity maps to rectify the input images prior to a direct image subtraction. An example application called TouchIt detects touch events between foreground objects (the user's hand for example) and the background surface. TouchIt uses two depth segmentation maps and simple logical operations to combine them into a touch map. We are able to generate 320x240 segmentation maps in 10ms on a 1GHz Intel PIII. This allows the TouchIt application to run in real time on an off-the-shelf PC.

Future Direction: Extend the algorithm to analytic surfaces and improve robustness.

Contact: Chris Wren
http://www.merl.com/projects/fast-depth-segmentation/

Lab: MERL Cambridge Research
Project Type: Research

Factorized Local Appearance Models



We propose a novel scheme for image-based object detection and localization by modeling the joint distribution of k-tuple salient point feature vectors which are factorized component-wise after an independent component analysis (ICA). Furthermore, we use a distance-sensitive histograming technique for capturing spatial dependencies which enable us to model non-rigid objects as well as distortions caused by articulation.

Background and Objectives: For appearance based object modeling in images, the choice of method is usually a trade-off determined by the nature of the application or the availability of computational resources. Existing object representation schemes provide models either for global features or for local features and their spatial relationships. With increased complexity, the latter provides higher modeling power and accuracy. Among various local appearance and structure models, there are those that assume rigidity of appearance and viewing angle, thus adopting more explicit models while others employ stochastic models and use probabilistic distance/matching metrics. Our objective is to model the high-order dependencies of local image structure by estimating the complete joint distribution of multiple salient point feature vectors using a density factorization approach.

Technical Discussion: We construct a probabilistic appearance model with an emphasis on the representation of non-rigid and approximate local image structures. We use joint histograms on k-tuples (k salient points) to enhance the modeling power for local dependency, while reducing the complexity by histogram factorization along the feature components. Although, the gain in modeling power of joint densities can increase the computational complexity, we propose histogram factorization based on independent component analysis to reduce the dimensionality dramatically, thus reducing the computation to a level that can be easily handled by today's personal computers. For modeling local structures, we use a distance-sensitive histograming technique. We argue in favor of collapsing the distance and instead using distance-dependent weights on the histogram increments. For example, for articulated and non-rigid object, any constraint on the structure or distance between distant points/regions can be misleading. In this case, inverse-distance-weighted histograming can be a better choice. A clear advantage of the proposed method is the flexibility in modeling spatial relationships. Experiments have yielded promising results on robust object localization in cluttered scenes as well as image retrieval.

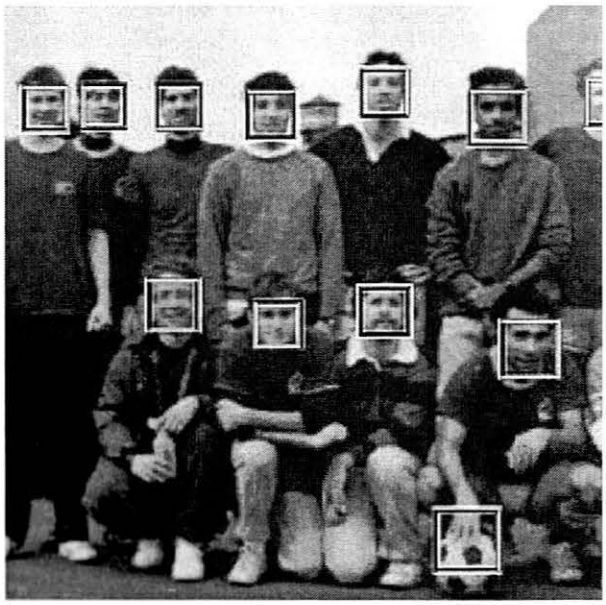
Collaboration: This project is joint collaboration Xiang Zhou and Thomas S. Huang (UIUC).

Future Direction: In the future, we plan to explore a more explicit way to incorporate spatial adjacency into the factorized local appearance model via graph matching.

Contact: Baback Moghaddam
http://www.merl.com/projects/flam/

Lab: MERL Cambridge Research
Project Type: Research

Fast Face Detection using a Cascade of Detectors



Automatic face detection is a critical component in the new domain of computer human observation and computer human interaction (HCI). There are many examples including: user-interfaces that can detect the presence and number of users; teleconference systems can automatically devote additional bandwidth to participant's faces; and video security systems can record facial images of individuals after unauthorized entry.

(Note: This work is the continuation of research begun before Viola joined MERL in March, 2001.)

Background and Objectives: In the past many other face detection approaches have been proposed - based on neural networks, support vector machines, and other types of machine learning. No existing approach has achieved 'real-time' processing. Our goal was to develop a highly reliable yet extremely efficient detector. Underlying this detector is a new machine learning approach that can be used for a variety of related tasks.

Technical Discussion: There are three main contributions of our object detection framework.

First: a new image representation called an Integral Image that allows for very fast feature evaluation. We use a set of features that are reminiscent of Haar Basis functions (though we will also use related filters which are more complex than Haar filters). Using the integral image, any one of these Haar-like features can be computed at any scale or location in constant time.

Second: a method for constructing a classifier by selecting a small number of important features using AdaBoost. In order to ensure fast classification, the learning process must exclude a large majority of the available features, and focus on a small set of critical features.

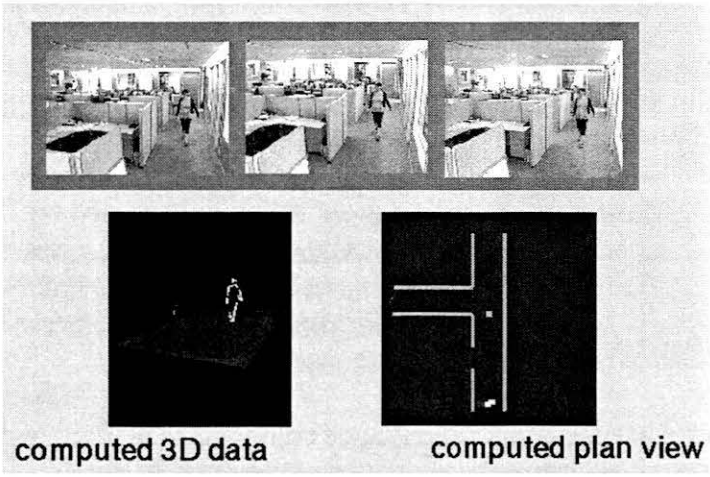
Third: a method for combining successively more complex classifiers in a cascade structure which dramatically increases the speed of the detector by focusing attention on promising regions of the image. The notion behind focus of attention approaches is that it is often possible to rapidly determine where in an image an object might occur. More complex processing is reserved only for these promising regions.

Future Direction: We are currently extending this approach to other real-time detection tasks such as pedestrian and car detection. We are also developing a related technology for real-time facial analysis.

Contact: Paul Viola
http://www.merl.com/projects/FaceDetection/

Lab: MERL Cambridge Research
Project Type: Research

Stereo Computer Vision for Observing People



See Color Figure 7

This project uses stereo computer vision to observe people. The primary goals are (a) to detect people in indoor and outdoor environments, (b) to do a 3D analysis of the position and motion of people e.g. to locate a person's position on a known map of the environment, or to compute statistics about the flow of people through an area (c) to understand the activity of people e.g. to classify standing-walking-running people. This technology is being developed in conjunction with other projects, on face detection and analysis, to create a framework which supports many applications of computer-human observation.

Background and Objectives: This work has many application areas including (a) surveillance - detecting intruders in a restricted area, (b) intelligent buildings - detecting people, estimating the number of people in an area, doing adult/child classification (c) intelligent traffic systems - estimating the size of crowds or lines in public transport hubs or at road-crossing places, detecting people in dangerous situations such as close to the edge of a train platform.

Technical Discussion: The major components of the system are feature detection, change detection, and stereo matching. Change-detection is based on features, rather than on pixels, as a superior approach with resilience to changing illumination. The stereo matching is currently based on a trinocular (3-camera) system, but is easily extendible to an arbitrary number of cameras. Cameras are becoming very cheap, so multi-camera stereo devices are likely to become much more common in the future.

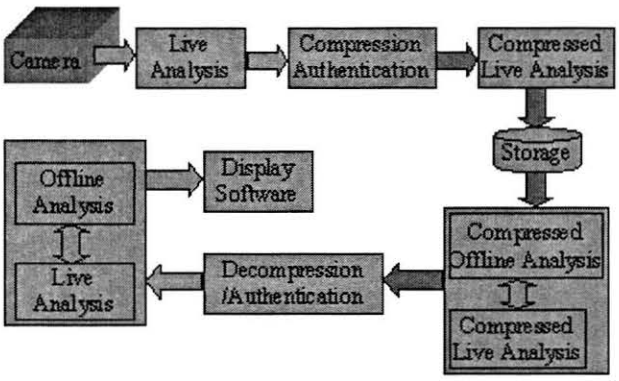
Collaboration: The feature detection software has been sent to Sanken.

Future Direction: The stereo demonstrations listed above, such as computation of 3D position, and measuring a person's height, are only initial ways of using the 3D data. Stereo converts ordinary images into images-plus-depth and the extra information offers many possibilities e.g. for 3D face modeling.

Contact: Paul Beardsley
http://www.merl.com/projects/CHO-stereo/

Lab: MERL Cambridge Research
Project Type: Advanced Development

Surveillance Architecture



Surveillance products are increasingly being differentiated by software. Future generations of surveillance systems will need to be flexible, scalable, robust, and will need to incorporate new software technologies easily, quickly, and effectively in order to compete. We are currently developing a modular software architecture that uses computer vision routines for surveillance. Our first components will use face detection, and stereo vision.

Background and Objectives: Melco has a presence in the surveillance components market, and the custom surveillance system design market. The computer vision technology and MPEG technology at MERL could help Melco increase its competitiveness in these areas. The object of the surveillance architecture project is to provide a robust, scalable, flexible, and modular architecture that adds value to Melco's next generation surveillance components, systems, and services.

Technical Discussion: MERL technologies relevant to surveillance include computer vision technologies such as face detection, face analysis, and stereo techniques, and MPEG technologies for video streaming, indexing, browsing, and searching. A software architecture for surveillance can provide a framework for integrating these technologies into Melco equipment and systems.

Future Direction: Moving forward we plan to develop modular computer vision components and an architecture specification to facilitate the incorporation of computer vision technologies in Melco products.

Contact: Joseph Woelfel
http://www.merl.com/projects/surveillance-architecture/

Lab: MERL Cambridge Systems
Project Type: Initial Investigation

Digital Communications

Digital Communications technology has revolutionized telecommunications and computer industries over the last decade. But technology development in this field is still rich with innovation. While mobile communication is being taken for granted today, many technological challenges remain, such as maintaining high quality of service in high mobility areas, maintaining high data transmission rates with constrained power and bandwidth, etc.

Two overarching questions drive the research efforts in Digital Communications at MERL. For wireless communications, the question is, "What services, network capabilities, and long-term architecture are needed to support applications beyond IMT 2000 and 3G?" For broadband communications, the question is, "What architectures and transmission technologies (optical, cable, wireless, powerline, etc.) are required to increase the bandwidth and services offered to the user?"

At MERL, we aim to create new concepts and technologies in wired and wireless Digital Communications. Participation in development of industry standards is an important part of our work. Our main focus is on three areas:

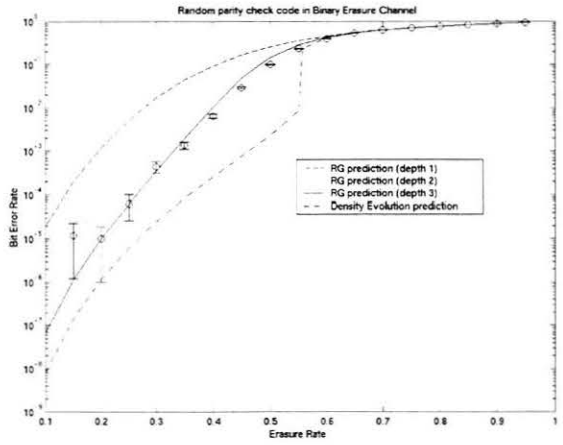
- Wireless Communications. MERL is an active participant in most of the major industry standards for wireless communications protocol, including IEEE 802.16, 3GPP2, 3GPP SG, ITU-T, and IETF. We are also applying these to standards improve system architecture and algorithm design for Mitsubishi Electric products. To test our designs, we have built a DSP-based test bed for FPGA prototyping.

- 10 Gigabit Ethernet. MERL is an active participant in the IEEE 802.3ae (10 Gigabit Ethernet) standard, and we are designing equalization techniques, building technology that will allow successful products based on the standard.

- Wireless LAN. MERL is an active participant in wireless LAN standards, including IEEE 802.11 and 1394 WLAN. We are also building test beds and demonstration systems, driving toward new technology for products and applications.

Project Descriptions

Analyzing and Designing Error-Correcting Codes



A fundamental problem in information theory is the analysis and design of optimal and efficiently decodable error-correcting codes of a given block-length N and rate r. The discovery of turbo-codes in 1993 and the re-discovery of Gallager codes soon thereafter stimulated considerable research into iterative decoding of codes on graphs. One can now argue that in the infinite block-length limit, we understand how to approach the Shannon limit. For practical applications, however, one must work with codes of finite block-length, for which existing theory is far from sufficient.

Background and Objectives: The best-known technique to analyze the performance of iterative belief propagation decoding is the "density evolution" method. However, this method presumes that the "messages" passed in the decoding algorithm are statistically independent, and is therefore only valid for the infinite block-length limit of certain codes such as irregular Gallager codes and their relatives. Our objective is to devise an efficient technique for computing the entire performance curve for an arbitrary parity-check code of finite block-length. This performance curve can then be used to obtain estimate the performance of a code at very low error-rates where simulation is impossible, or as an objective function in a search over codes.

Technical Discussion: Our technique is based on the "renormalization-group" approach from physics: the idea is to continually replace an error-correcting code with a simpler error-correcting code that has nearly identical performance, until the code is reduced to a small enough size that its performance can be computed exactly.

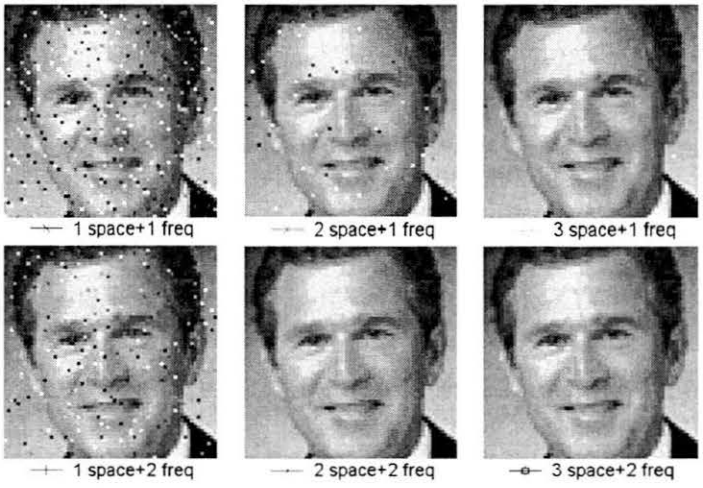
Collaboration: With Jean-Philippe Bouchaud (CEA-Saclay) and Erik Sudderth (MIT).

Future Direction: We are exploring improved techniques which promise to better handle codes whose graphical representations have short loops, and anticipate actually searching for optimal finite block-length codes using our techniques.

Contact: Jonathan Yedidia
http://www.merl.com/projects/ECC/

Lab: MERL Cambridge Research
Project Type: Research

STF Diversity Combining for Wireless Communication



MERL Murray Hill developed an advanced multi-user receiver for the next generation wireless communication systems. In order to maintain good communication quality in high mobility and high data rate transmission environments, this receiver utilizes joint space-time-frequency diversity combining technique and adaptive interference cancellation techniques to mitigate multiple access interference (MAI) and channel impairments in Code Division Multiplex Access (CDMA) communication systems. The performance improvement, in terms of error rate and system capacity, over the conventional receiver can be shown through the computer simulations.

Background and Objectives: CDMA technology is the key technology in the 3rd generation wireless communication systems. Multi-user detection is an important issue in CDMA systems since each user shares the same frequency band and time interval. In order to support multimedia communication in the future, a high performance CDMA receiver with high data rate transmission in high mobility environments is desired.

Technical Discussion: Diversity techniques are used in practice to combat fading channels. The increased mobility of users in cellular communication often results in fast fading or large Doppler spreads and the degradation of system performance is expected. The use of joint time-frequency diversity techniques provides significant performance improvement over the existing systems in the single user receiver design. Unlike the time-frequency diversity is the wireless channel characteristics, one can create diversity dimension artificially, such as space-time diversity for multi-user detection in CDMA systems. Diversity gains can be achieved by using multiple elements for antenna arrays processing and adaptively combining space-time-frequency diversity signals. However, due to heavy computational complexity, the trade off between performance and complexity become an importance issue when applying this technique.

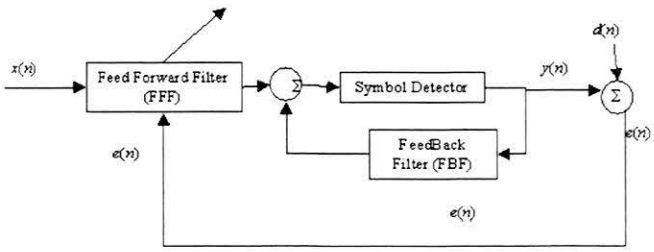
Collaboration: MHL collaborates with Princeton University for the development of diversity combining techniques for future wireless communication systems.

Future Direction: Trade off between complexity and performance.

Contact: Jyhchau "Henry" Horng
http://www.merl.com/projects/diversity-technique/

Lab: MERL Cambridge Research
Project Type: Advanced Development

IEEE 802.3ae Equalizer Project



MELCO is currently developing a 10Gigabit per second optical communications transceiver module. This transceiver is based on the new IEEE 802.3 Ethernet standard that is currently under development.

Adaptive equalization technology has been identified for possible incorporation into MELCO transceiver product. Equalizers can be used to overcome dispersion introduced by the optical fiber and other system components and improve system performance as well as increase transmission distance. It also a key technology that allows MELCO to differentiate it's products from many other 10Gigabit Ethernet suppliers. MERL-MH is developing adaptive equalizer algorithms and assisting MELCO with the implementation of the equalizer.

Background and Objectives: The IEEE 802.3 Working Group develops standards and specifications for Carrier Sense multiple Access with Collision Detection (CSMA/CD) (commonly known as Ethernet) based LANs. The main goals of the working group are extend the Ethernet operating speed to 10 Gb/s and to expand the Ethernet application space to include longer distance links (>10 km).

Technical Discussion: While adaptive equalizers have been used in many communication systems in the past, there use in optical data communications is still relatively new. The challenge here is to develop equalizers that deliver improvement over a wide range of channel conditions and multiple types of physical media. The new Ethernet standard defines both multimode fiber and single mode fiber for use as physical media. In each case the fiber type determines the dispersion mechanisms, differential mode dispersion in multimode fiber and chromatic dispersion in single mode fiber. The dispersion of the transmitted pulses results in inter-symbol interference (ISI) and increased bit error rate at the receiver. In addition, the ISI eventually limits the distance that signals can propagate and still be recovered without error.

Preliminary results have shown that adaptive equalizers are capable of compensating for the ISI expected from today's installed based of optical fiber. Currently, it is believed that an equalized version of 10 Gigabit Ethernet can achieve more than twice the transmission distance then the basic version without equalization technology.

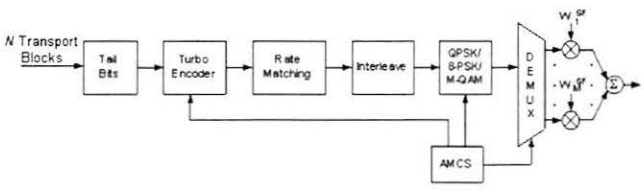
Collaboration: MERL-MH is actively collaborating with L-ji-se System Development Group.

Future Direction: Future work will focus on integration of proposed equalizers with MELCO transceiver design and improving equalizer performance.

Contact: Philip Orlik, Daqing Gu
http://www.merl.com/projects/10GbEqualizer/

Lab: MERL Murray Hill
Project Type: Advanced Development

HSDPA in W-CDMA



HSDPA (high speed downlink packet access) is the key features in the 3rd generation wireless communication standard W-CDMA (wideband CDMA), which will be launched in late 2001 by NTT DoCoMo in Japan. This technology provides high data rate transmission in CDMA downlink to support multimedia services. We researched and developed the key technologies on HSDPA, such as AMC, HARQ, and MIMO, to enhance MELCO's IPRs in this area and prepare the essential technologies for 4th generation wireless communication.

CDMA: code division multiple access, AMC: adaptive modulation & coding, HARQ: hybrid automatic-repeat-request, MIMO: multiple input multiple output, 3GPP: 3rd generation partnership project. QoS: quality of service, UE: user equipment.

Background and Objectives: The 3G standard WCDMA continues to evolve based on research and new innovations. Release 4 specifications will provide efficient IP support enabling provision of services through an all-IP core network. The later 3GPP releases, HSDPA, will provide even higher data rates up to approximately 10 Mbps to support multimedia services (packet-based) in the future. The objective of this project is to identify the key features of HSDPA and develop the key technologies for HSDPA to enhance MELCO's IPR portfolio on 3G systems.

Technical Discussion: HSDPA provides high data rate transmission up to approximately 10 Mbps over a 5MHz bandwidth. To reach this target, some key technologies are used for HSDPA: AMC, HARQ, MIMO antenna processing, fast cell search, stand-alone dedicated share channels, and downlink channel structure. In general, all HSDPA users share the channel in both time and code domains. Adaptive modulation and coding is used to support multiple rate transmission for different types of multimedia services. Some other features are used to enhance the system performance, such as system throughput increase, QoS requirement, interference cancellation, and high data rate transmission. Due the high rate transmission, the complexity and capability of UE and base-station become an important issue and therefore, the trade off between complexity and performance should be considered.

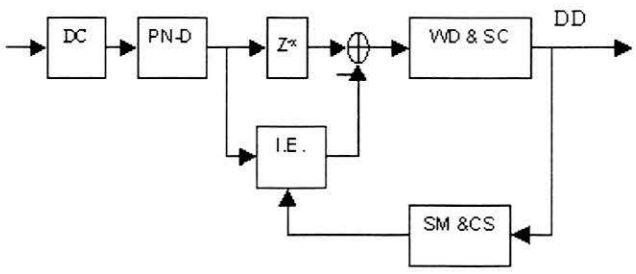
Collaboration: MHL collaborates with Y-Kenn of Tsu-Kai-se for the development of HSDPA in WCDMA.

Future Direction: Integration of WLAN with cellular systems.

Contact: Jyhchau "Henry" Horng
http://www.merl.com/projects/HSDPA/

Lab: MERL Murray Hill
Project Type: Advanced Development

Interference Cancellation in Next Generation CDMA Systems



The adaptive interference cancellation (IC) algorithm has been developed for the next generation CDMA systems. It can be used for a low rate user to cancel the interferences from high rate users. An interference estimate, IE, is designed to estimate the high-rate user's signal (interference). The estimated interference is then subtracted from the signal before the Walsh demodulation and soft combining. The interference cancellation improves the SNR, and thus the system performance. The SNR measurement & control system block, SM & CS, calculates the SNR and triggers the interference cancellation when it is necessary.

Background and Objectives: Variable transmission rate services will be supported simultaneously in the next generation wireless systems. In practice, the power control is used to meet the predefined performance for different users. In an additive white Gaussian noise channel, un-even power is then needed for different rate users to obtain the same bit-error-rate (BER). However, when the power difference between the different rate users becomes large, the performance of the low rate user will be seriously degraded. The adaptive IC is to address this issue by estimate and canceling the interference from the high rate user.

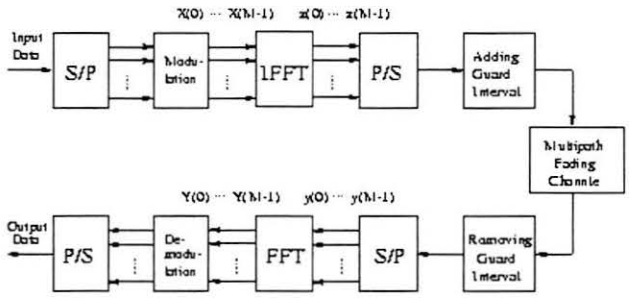
Technical Discussion: By changing the spreading gain for each user, variable rate transmission can be achieved. Data services generally need much lower BER than the voice services, therefore higher power. The power ratio between high rate users and low rate users will be further enlarged when two users are close to each other in location. To maintain a good conversation quality for the voice service while another data service user access to the network with high data rate, becomes very important issue in the CDMA system downlink.

Collaboration: We have developed the above technique in the past year, 1 paper has been published and 1 patent is applied. Currently we are working with YRP and Joho-soken for possible collaborations on 3G/4G wireless Systems design topics.

Future Direction: The interference cancellation is one of the key techniques for improving the service performance of a wideband CDMA system. We will continue our interference cancellation study, and explore if we can combine this technique with other techniques together to achieve better QoS for the future wideband CDMA systems.

Contact: Jin Zhang **Lab:** MERL Murray Hill
http://www.merl.com/projects/interference-cancellation/ **Project Type:** Advanced Development

OFDM Communication Systems



Orthogonal Frequency Division Multiplexing (OFDM) is a promising technique for transmitting high bit rate data over wireless communication systems. It has been adopted in many wireless communication standards, such as Digital Audio Broadcast (DAB), wireless local area network (WLAN) and Hyper LAN, and is a potential candidate for 4th generation wireless communication systems. MH Lab has developed the key techniques in OFDM system, such as synchronization and channel estimation. By using the developed techniques, system performance can be improved for high data rate transmission.

Background and Objectives: In mobile wireless communications, intersymbol interference due to fading and multipath impairments causes error in received signals. To combat these impairments, Orthogonal Frequency Division Multiplexing (OFDM) is considered to be a solution. OFDM is known to be robust against multipath propagation and channel noise. However, in real system design, there are still some factors, which will cause the performance degradation in OFDM systems, such as synchronization problems and channel estimation problems. These problems become even more critical in the future wireless communication systems for high data rate transmission.

Technical Discussion: OFDM is a multicarrier transmission. The basic idea of OFDM is to divide the available spectrum into several subchannels (subcarriers). Almost flat fading can be achieved in each subchannel by making all subchannels narrowband, which makes equalization very simple. To obtain a high spectral efficiency the frequency response of the subchannels are overlapping and orthogonal. This orthogonality can be completely destroyed if frequency offset between transmitted and received signal is big enough. In addition, for high data rate transmission, fast data frame acquisition is desired. Synchronization problems therefore are very important in OFDM systems. Furthermore, in the absence of channel information, differential detection is usually used at the expense of 3 to 4 dB loss in signal to noise ratio (SNR) compared to coherent detection. To make coherent detection applicable, an efficient channel estimation technique is called for.

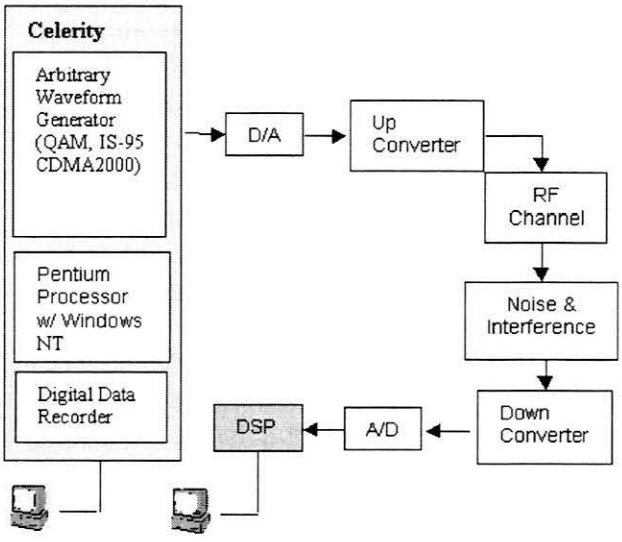
Collaboration: MHL collaborates with Princeton University for the development of channel estimation techniques in OFDM systems.

Future Direction: We are currently developing simulation models on Simulink as well as real-time DSP system models.

Contact: Jyhchau "Henry" Horng, Charles You
http://www.merl.com/projects/OFDM/

Lab: MERL Murray Hill
Project Type: Advanced Development

Real-time DSP-based Testbed



The Murray Hill Laboratory of Mitsubishi Electric Research Laboratories set up a real-time DSP-based test platform. The main features of the testbed are its programmability and capability of real-time high data rate processing. The signal generator is programmable and capable of generating any type of standards or signal types in real-time; The DSP board combined with a digital receiver can receive the high frequency wideband or narrowband analog signals and downconvert it to the IF and convert the analog signal to digital samples. Then the digital samples are processed through DSPs and the outputs are generated. The testbed will be used for demonstrating WCDMA and 4G wireless systems.

Background and Objectives: Though 3G wireless systems will be deployed into market very soon, there is still research to be done in the 3G . More importantly, the testbed will be used for developing 4G systems. Therefore, the testbed is essential for the real-time development and testing for future wireless systems.

Technical Discussion: Celerity system includes a dual channel analog arbitrary waveform generator module and a single channel analog data acquisition module. Each module is configured with 256 MB of data memory, with a maximum I/O rate of 300 MB/s. The AWG module is configured with a dual channel 12 bit 75 MS/s analog output daughtercard which can provide coherent dual channel baseband outputs at rates up to 75 MS/s per channel. The Acquisition Module is configured with a single channel 12 bit 65 MS/s analog input daughtercard. The wideband digital receiver has two input channels including amplifier, A/D and programmable downconverter. Each channel input can receive bandwidths up to 30 MHz and has a 12-bit A/D converter with maximum sampling rate up to 65 MHz. The output of the downconverter delivers I and Q complex outputs to the mezzanine FIFO of the 'C6201 DSP board. The wideband digital receiver is attached directly on top of the Quad 'C6201 DSP board via VIMbus. TI TMS320C6201 processor is a fixed point with a clock speed of 200 MHz.

Future Direction: CDMA2000 & WCDMA demonstration.

Contact: Charles You
http://www.merl.com/projects/CDMA-testbed/

Lab: MERL Murray Hill
Project Type: Advanced Development

Graphics

Most of our sensory information comes through our eyes. Computer Graphics, therefore, plays a critical role at the interface between human and computer.

MERL has a strong history in the field of computer graphics research, with participation in respected industry conferences and with diverse research projects, running the gamut from scientific visualization to interactive manipulation of geometric shapes.

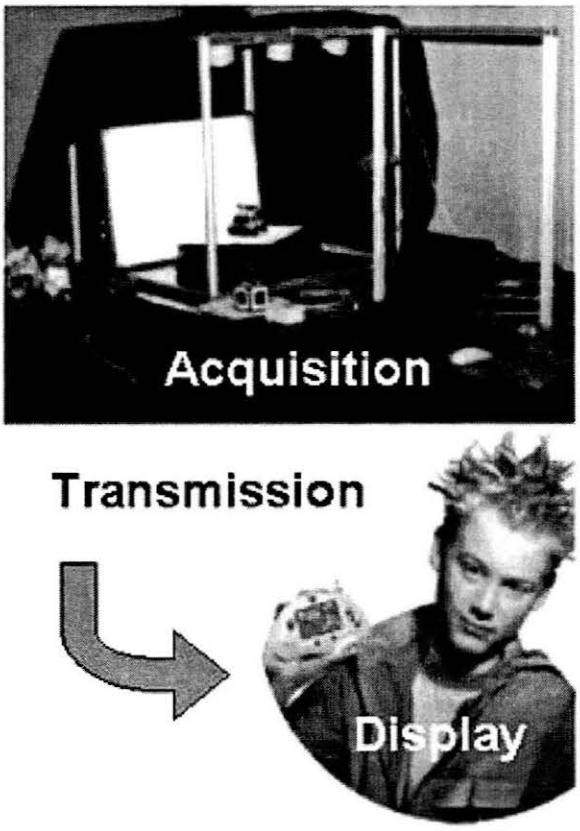
At MERL, our key focus areas in computer graphics include:

- Modeling. We are developing new methods of representing computer graphics that offer advantages for compression, transmission, editing, rendering, and animation of graphics models with increasingly complex shapes. Our current research includes the use of Adaptively Sampled Distance Fields, a new digital representation of shape applied to several projects in this section.

- Image-Based Rendering (IBR). IBR refers loosely to techniques that generate new images from other images rather than from a geometric description. It combines techniques of computer graphics and computer vision. IBR enables interactive photo-realistic rendering, because the source data are photos. We are working on the creation of static and dynamic 3D photographs and immersive 3D environments, enabling new applications in e-commerce, entertainment, and tele-collaboration. This is described in detail in the 3D Images project on page 90.

Project Descriptions

3D Images



3D Images is an ongoing research project at MERL with the goal to develop core technology to acquire, transmit, and display complex, real-life three dimensional objects as efficiently as possible. A 3D Image is the extension of a 2D image into three dimensions: it is a point sample representation that allows the user to display a 3D object from arbitrary viewpoints.

The fundamental approach of this research is to use surfels (or surface elements) as an alternative display primitive to polygons [1]. 3D image objects are represented as a dense set of surface point samples. 3D images offer complex shape, low rendering cost, progressive rendering, and high image quality, which makes them specifically suited for low-cost, real-time graphics, such as portable graphics on cell phones or PDAs. Potential markets include eCommerce and 3D games.

Background and Objectives: We see new opportunities for 3D graphics in today's changing technology landscape. PC graphics has become a commodity, and the internet and wireless devices have created new opportunities. The core objective of this project is to develop 3D rendering technology for wireless and eCommerce applications. A second objective is to develop 3D scanning technology for real objects. We have applied for 9 patents.

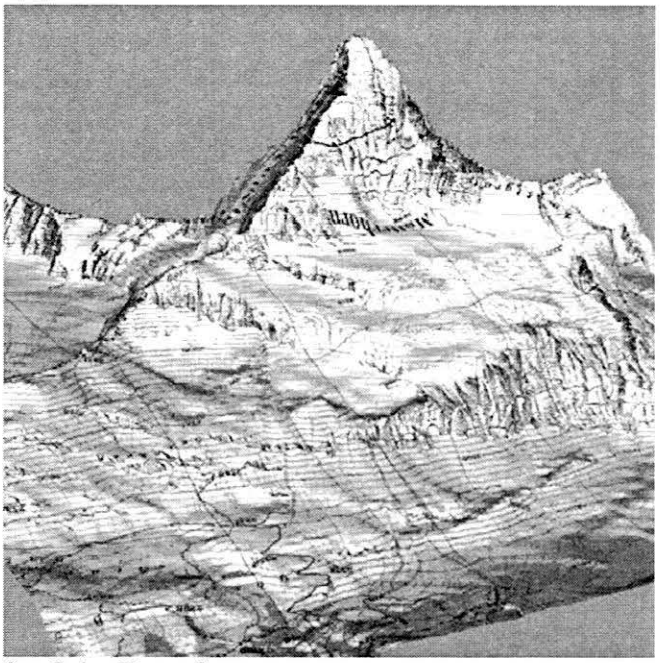
Technical Discussion: We have developed technology to acquire 3D images using a turntable-based system with a video camera. The object is automatically rotated on the turntable and a number of images are taken. From the 2D images we reconstruct a 3D image using an image-based visual hull algorithm. We also have a complete rendering system in software to display 3D images as described in [1]. It is an efficient implementation that displays 3D images at interactive frame rates on a 700 MHz Pentium III processor.

Future Direction: We are working on compression and progressive transmission of 3D images while improving the rendering pipeline.

Contact: Hanspeter Pfister
http://www.merl.com/projects/3Dimages/

Lab: MERL Cambridge Research
Project Type: Research

EWA Splatting



See Color Figure 8

Elliptical Weighted Average (EWA) splatting is a novel technique for high quality rendering of point-based surface and volumetric objects. We introduce screen space EWA filtering, which provides anisotropic texture filtering. Moreover, EWA splatting can be used to render surface and volume data in the same rendering pipeline.

Background and Objectives: Modern acquisition devices, such as laser range or MRI scanners, have produced some of the most complex models to date. Because of their huge volume of data, these models are extremely challenging for post-processing or rendering algorithms. A commonly used approach is generating triangle meshes from the point data and using mesh reduction techniques to render them. However, some scanned meshes are too large to be rendered interactively, and some applications cannot tolerate the inherent loss in geometric accuracy and texture fidelity that comes from polygon reduction. Moreover, complex 3D objects are rendered with triangles that often are approximately the size of a pixel in conventional rendering pipelines. Rasterizing these tiny primitives leads to inefficient operation of triangle based rendering pipelines.

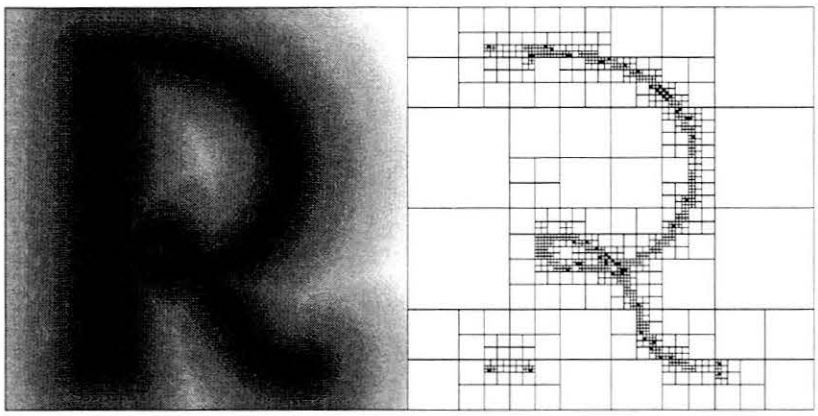
Technical Discussion: Instead of using triangles, we propose a novel point-based rendering algorithm called EWA splatting, focusing on high quality texture filtering. We introduce a novel screen space formulation of the EWA filter, the best anisotropic texture filtering algorithm for interactive systems. Our novel splat primitive provides superior image quality compared to previous splatting techniques. This makes EWA splatting applicable to high-resolution laser range scans, terrain with high texture detail, volume data, or point-sampled geometric objects. EWA splatting can also be used to mix surface and volume data in the same framework. We have applied for 1 patent.

Future Direction: We are working on extending the EWA splatting approach to irregularly sampled volume data.

Contact: Hanspeter Pfister
http://www.merl.com/projects/ewa-splatting/

Lab: MERL Cambridge Research
Project Type: Research

Adaptively Sampled Distance Fields (ADFs)



ADFs are a new digital representation of shape with several advantages over existing approaches. They provide efficient and accurate representation of both smoothly curved surfaces and surfaces with fine detail. ADFs have the potential to impact many diverse industries including CAD/CAM (simulation, path planning, and verification for milling precision parts), Entertainment (building models for games and movies), Fonts (high-quality display of letterforms for PDAs), 3D Scanning (3D models from image or range data), 3D Printing (rapid prototyping), and Color Management (projectors, PDAs, monitors, and printers).

Background and Objectives: Our objectives include fundamental research, incorporation of this research into a product-worthy C library ready for commercialization, development of a comprehensive patent portfolio, and collaboration with key industrial players to refine and expand the vision for ADFs.

Technical Discussion: A distance field is a scalar field that specifies a distance to a shape, where the distance may be signed to distinguish between the inside and outside of the shape. ADFs consist of adaptively sampled distance values, organized in a spatial data structure, with a method for reconstructing the distance field from the sampled distance values. This approach permits the accurate and compact representation of fine detail and smooth surfaces, together with efficient processing. ADFs allow: the representation of more than the surface (interiors and exteriors); the compact representation of sharp features and organic shapes; smooth surface reconstruction; trivial inside/outside and proximity testing; fast and simple CSG operations; fast geometric queries such as closest point; and efficient computation of surface offsetting, blending and filleting, collision detection, morphing, and rough cutting.

Contact: Sarah Frisken , Thouis "Ray" Jones, Ron Perry **Lab:** MERL Cambridge Research
http://www.merl.com/projects/adfs/ **Project Type:** Research

Kizamu: A System For Sculpting Digital Characters

This report presents Kizamu, a computer-based sculpting system for creating digital characters for the entertainment industry. Kizamu incorporates a blend of new algorithms, significant technical advances, and novel user interaction paradigms into a system that is both powerful and unique.

At the heart of the Kizamu system are Adaptively Sampled Distance Fields (ADFs), a volumetric shape representation with the characteristics required for digital clay. In this report, we describe the system and present the major research advances in ADFs that were required to make Kizamu a reality.

Background and Objectives: To meet the demands of high-end digital character design, Kizamu addresses three requirements posed to us by a major production studio. First, animators and artists want digital clay - a medium with the characteristics of real clay and the advantages of being digital. Second, the system should run on standard hardware at interactive rates. Finally, the system must accept and generate standard 3D representations thereby enabling integration into an existing animation production pipeline.

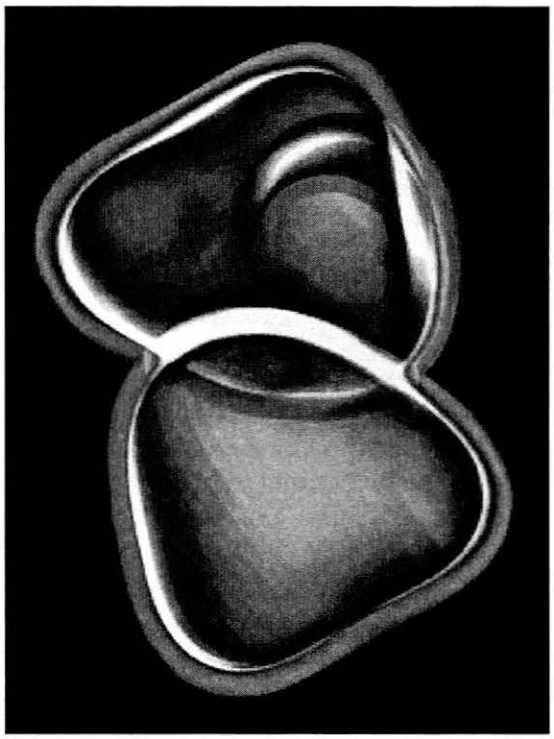
Technical Discussion: The contributions of the Kizamu system fall into 4 major areas: 1. Innovations in the volumetric sculpting interface that take advantage of the distance field to provide more control to the artist such as using the distance field to orient and position the sculpting tool relative to the surface, using distance values to constrain sculpting tools, and region-based conversion to triangle patches for editing via control vertices; 2. Advances in ADF generation and editing such as a new generation algorithm with reduced memory requirements, better memory coherency, and reduced computation, as well as a method for correcting distance values far from the surface during sculpting; 3. Several new ADF rendering approaches that take advantage of hardware acceleration in standard PCs, including fast generation of triangles and surface points; 4. Methods for inputting and outputting models from the system including an improved technique for generating ADFs from scanned range images and a very fast new algorithm for generating topologically consistent (i.e., orientable and closed) triangle models from ADFs at various levels of detail (LOD).

Future Direction: We are integrating the new algorithms into a product-worthy C library for ADFs.

Contact: Ron Perry, Sarah Frisken
http://www.merl.com/projects/ADF-Kizamu/

Lab: MERL Cambridge Research
Project Type: Research

A New Framework For Non-Photorealistic Rendering



See Color Figure 9

Non-photorealistic rendering, or NPR, has emerged as an important field of computer graphics. Most NPR methods attempt to create imagery mimicking a particular style produced by an artist. Several such styles have been investigated, including painting, watercolor, pen and ink, charcoal, and cartoon coloring. Thus far, most published NPR algorithms focus on a specific artistic style. Underlying these diverse artistic effects, however, are several recurring themes common to most NPR techniques.

We have developed a novel framework for NPR, based on adaptively sampled distance fields (ADFs), which uses these recurring themes to enable many diverse styles to be realized.

Background and Objectives: Non-photorealistic rendering is capable of broadening our ability to communicate thoughts, emotions, and feelings through computers. Artists have learned that by making images look less photorealistic they enable audiences to feel more immersed in a story. NPR permits the use of real-time stylized rendering to create a more compelling storytelling experience, whether that be through dynamic images in an on-line book, interactive technical illustrations, or immersive cartoon experiences.

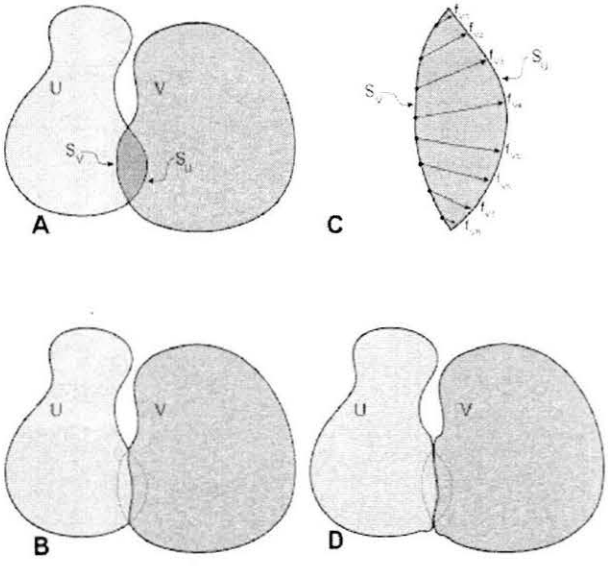
Technical Discussion: By representing a model as an ADF, we can interactively and accurately generate view-dependent particles (for stroking and coloring) and view-dependent triangles (for stroking, coloring, and visibility determination). From these view-dependent elements, many diverse styles can be realized by employing existing techniques. The multi-resolution nature of ADFs provide regulation of stroke (particle and triangle) density, guaranteed frame rates, and optimal use of processing resources in a system that scales both to new hardware and to models of increasing complexity. This approach thus unifies several operations common in artistic rendering. Furthermore, because ADFs can represent both hard surfaces and soft organic volumetric forms, opportunities exist to develop new volumetric NPR techniques in this single framework. We have developed such a volumetric style depicted above. Finally, the distance field provides other advantages, such as providing new ways for generating orientation fields, which can be exploited in existing techniques.

Future Direction: To continue fundamental research and to embody this research in a product-worthy C library.

Contact: Ron Perry, Sarah Frisken
http://www.merl.com/projects/ADF-NonPhotoRendering/

Lab: MERL Cambridge Research
Project Type: Research

A Computationally Efficient Framework for Modeling Soft Body Impact



Any system that models impacts between soft bodies must be able to detect collisions, compute impact forces, and provide an initial estimate of contact deformation. Adaptively Sampled Distance Fields (ADFs) provide a computationally efficient framework that enables these operations to be performed quickly, with efficient use of memory, and more accurately than previous methods.

Important application areas include: simulation of non-rigid structures in computer-aided design; animation of cloth, skin, and organic material in the entertainment industry; and surgical planning, simulation, and intra-operative guidance in the medical industry.

Background and Objectives: While there has been significant progress in simulating collisions between rigid bodies, much remains to be done for modeling interactions between soft bodies. Graphical techniques for representing and deforming soft bodies range from non-physical (e.g., control point-based) to physically plausible (e.g., Free Form Deformation - FFD) to physically realistic (e.g., Finite Element Modeling - FEM). All of these techniques require three operations to model interactions between soft bodies: 1) detecting collisions between deforming bodies, 2) computing impact forces when bodies collide, and 3) determining deformation forces or contact deformation of the bodies to initialize a deformation technique such as one of those listed above.

Technical Discussion: This work exploits certain properties of ADFs (e.g., compact surface representation, trivial inside/outside and proximity tests, fast localization of potential contact regions, more accurate representation of the overlap region, and simple methods for computing material-dependent contact deformation) to perform these three operations efficiently.

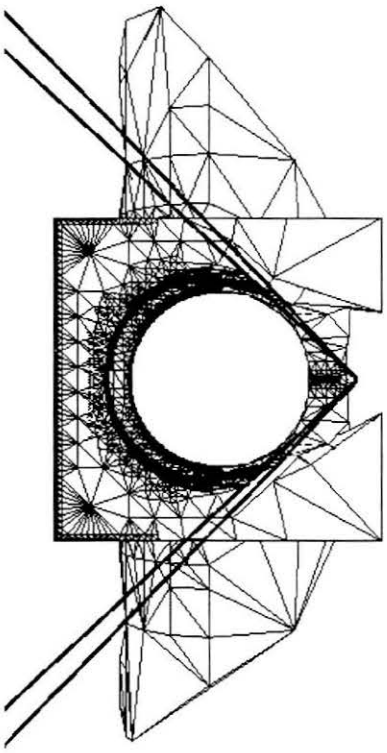
When detecting collisions between two ADFs, their spatial data structures can be exploited to quickly localize potential regions of overlap. Within these regions, a new ADF of the shape defined by the intersection of the two ADFs is locally generated. If the intersection ADF is non-empty, a collision is detected and penalty-based methods are used to compute the impact forces, making use of the ADF distance values and the gradient of the distance field. The implicit nature of the distance field is used to estimate an initial contact deformation that can then be used in a physically-based deformation technique (e.g., FEM).

Future Direction: We are continuing fundamental research and exploring the application of ADFs to physical simulation.

Contact: Ron Perry, Sarah Frisken
http://www.merl.com/projects/ADF-SoftBody/

Lab: MERL Cambridge Research
Project Type: Research

Dynamic Meshing Using Adaptively Sampled Distance Fields



We have developed a new method for generating viewpoint-dependent dynamic triangle meshes using ADFs (Adaptively Sampled Distance Fields). The technique produces detail-directed triangle meshes of high visual quality as viewed from the camera, while minimizing the number of triangles in non-visible portions of the object. It meets frame rate criteria (currently at 30 frames per second it maintains ~25K triangles), even during viewpoint changes that lead to large differences in the visible portion of the object.

The method has applications in several areas of the entertainment industry, including games and non-photorealistic rendering, where rendering the large numbers of triangles required to represent complex models limits interactivity.

Background and Objectives: Many models used in real-time graphics applications are generated automatically using techniques such as laser-range scanning. The resultant meshes typically contain one or more orders of magnitude more polygons than can be displayed by today's graphics hardware. Numerous methods have been proposed for automatically creating level-of-detail (LOD) meshes from large input meshes. These techniques typically generate either one or more static LOD meshes, pre-computed before use in the application, or a dynamic mesh, where the LOD of the mesh adapts to frame rate requirements. We have developed a new dynamic LOD technique ideal for applications such as games and physical simulations based upon ADFs.

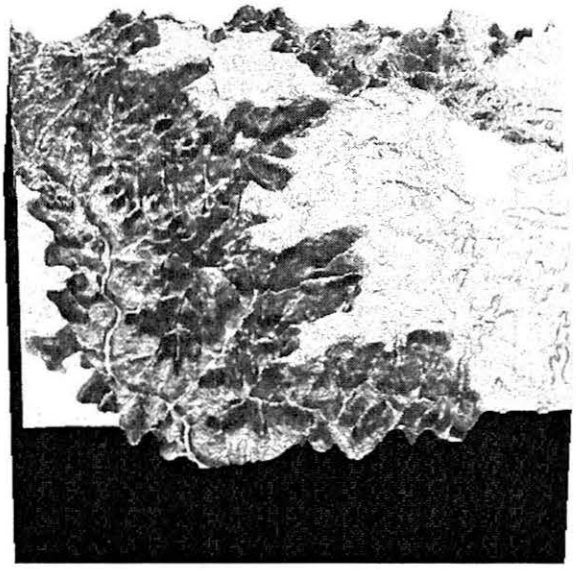
Technical Discussion: ADFs are a new shape representation which adaptively sample the signed distance field of an object and store the distance values in a spatial hierarchy (we use an octree for dynamic meshing). We utilize a fast, new triangulation method that generates topologically consistent (orientable and closed) triangle meshes from the ADF structure. The technique exploits the hierarchical nature of the octree to produce detail-directed triangles.

Our method creates a triangle mesh from the ADF, associating triangles with ADF cells, and then adapts the mesh in real-time to current viewing parameters in order to optimize visual quality while satisfying a user-defined frame rate. The algorithm is composed of two stages: a pre-processing stage and a real-time stage. The pre-processing stage initializes data for the real-time stage and creates an initial, view-independent active cell list from which a triangle mesh is derived. The real-time stage adapts and optimizes the active cell list and corresponding triangle mesh for the current viewing conditions.

Future Direction: We are exploring the application of this method to the entertainment industry for computer games and non-photorealistic rendering.

Contact: Sarah Frisken , Ron Perry
Lab: MERL Cambridge Research
http://www.merl.com/projects/ADF-DynamicMeshing/
Project Type: Research

Computing 3D Geometry Directly from Range Images



3D range sensing has proven to be an invaluable technique for generating models for the entertainment industry. We have developed a new method for computing 3D geometry from range data that: 1) computes distances directly from range images rather than from range surfaces, 2) generates an Adaptively Sampled Distance Field (ADF) rather than a distance volume or 3-color octree, resulting in a significant savings in memory and distance computations, 3) provides an intuitive interface for manually correcting the generated ADF, and 4) generates optimal triangle models from the ADF (with fewer triangles in flat regions and more triangles where needed to represent surface detail) using a fast new triangulation method.

Background and Objectives: Several techniques have been developed in research and industry for computing 3D geometry from sets of aligned range images. Recent work has shown that volumetric methods are robust to scanner noise and alignment uncertainty and provide good quality, water-tight models. However, these methods suffer from limited resolution, large memory requirements and long processing times, and they produce excessively large triangle models. By making use of ADFs, we can reduce the memory and distance computation requirements, edit occluded regions directly in the ADF, and produce optimal triangle models.

Technical Discussion: Constructing 3D range surfaces and computing distances from these surfaces contribute significantly to the computational requirements of volumetric methods. If, instead, the distance field could be generated directly from 2D range images, model generation times could be reduced. However, range images do not provide true distance data. We have developed a method to correct the 3D projected distance field that divides sampled distances by the local gradient magnitude, resulting in a better approximation of the true distance field near the surface and better results when combining projected distance fields. Since computing the local 3D gradient to make this correction could be prohibitive (it requires 6 additional distance computations), we have also developed a method to derive the 3D gradient from a 2D gradient image generated once during preprocessing, resulting in significantly faster generation.

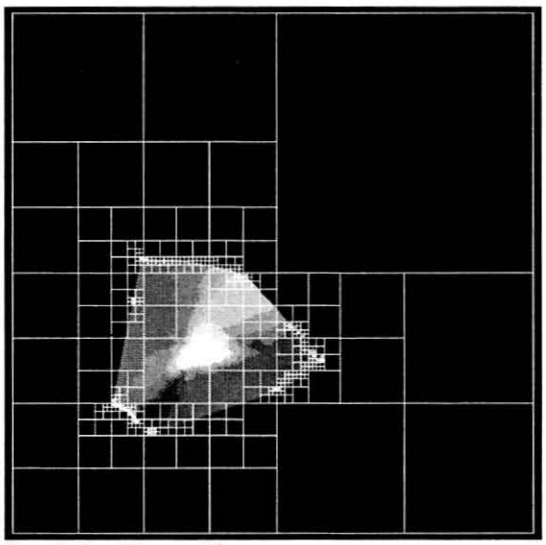
In addition, we have developed a fast algorithm for triangulating the ADF octree that generates an optimal triangle mesh and can produce LOD (Level-of-Detail) models ideal for applications such as games and physical simulations.

Future Direction: We are integrating this method into a product-worthy C library for ADFs.

Contact: Sarah Frisken , Ron Perry
http://www.merl.com/projects/ADF-3DScanning/

Lab: MERL Cambridge Research
Project Type: Research

A New Representation for Device Color Gamuts



See Color Figure 10

An important part of color reproduction is the rendering of colors onto display devices, where the colors may be spot colors (e.g., tints) or the colors of a digitized photograph. Display devices, such as monitors and projectors, are capable of showing only a subset of all perceivable colors; this subset is called the device color gamut.

We have identified a new variation of adaptively sampled distance fields (ADFs) for representing device color gamuts that overcome limitations of previous work. Associated algorithms for gamut testing and gamut construction have been developed.

Background and Objectives: Colors subjected to a reproduction system may not fall within the gamut of the target display device. In these cases, the colors cannot be exactly reproduced. If no corrective action is taken, the reproduction of graphics having out of gamut colors results in undesirable distortions of the tonal and color range of the graphic. When this occurs, some action is required, either notifying the user to modify the colors to fit within the gamut, or employing an algorithm to automatically compress the colors according to some predetermined strategy to fit within the gamut. In either case, a reliable technique is needed to determine whether or not a color lies inside a device gamut. When considering applications involving high resolution digitized images containing millions of colors, this gamut test must be very fast.

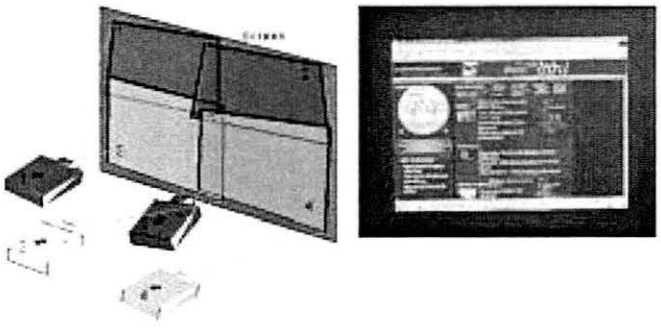
Technical Discussion: The octree ADFs (and algorithms) previously suggested are not optimized for gamut testing; the ADFs consume too much memory, lack memory coherency, take too long to generate, and are limited (in practice) to a maximum tree depth of 9. To overcome these problems, we construct a contiguous ADF representing the gamut using a new tiled bounded surface generation algorithm. Representing the ADF in a contiguous block is important for conforming to the ISO ICC computational framework (the international standard for color reproduction adopted by the industry). We have also trimmed the cell data structure to contain only the necessary elements for reconstruction (the primary operation in gamut testing) and have found that 8 or 16 bit distance values (rather than 32) are sufficient. This new ADF representation is compact and amenable to a hardware implementation.

Collaboration: MELCO develops digital projectors requiring color management, thus providing a potential collaborator for this research.

Future Direction: To continue fundamental research and to embody this research in a product-worthy C library.

Contact: Ron Perry, Sarah Frisken
Lab: MERL Cambridge Research
http://www.merl.com/projects/ADF-ColorGamut/
Project Type: Research

Low Cost Projector Mosaic



We have developed a set of techniques to create large format displays using casually installed overlapping projectors. The image alignment and intensity blending is completely automatic. The time required to calibrate the system is less than 20 seconds allowing a very easy to use and flexible setup. Current multi-projector systems cost several times the price of the projectors due to the expensive infrastructure and maintenance. Our techniques focus on software methods to reduce the cost by eliminating the need of rigid support structures and manual alignment.

Background and Objectives: A photo-mosaic is a collection of images registered together to form one large image. Can we create a similar projector mosaic on a display surface by seamlessly merging output of overlapping projectors? For photo-mosaic, the images are captured by casually orienting and positioning the camera. Can we similarly create a large display using casually aligned projectors? Currently, large displays are generated by tiling together a two dimensional array of projectors by precisely aligning them. Due to such design constraints, the installation and operation of such systems is extremely expensive. If we allow casual approximate installation, the cost of multi-projector systems can be greatly reduced. Further, if the registration and blending of overlapping images is performed completely automatically, such systems can become very easy to use allowing wide spread use in homes, offices and shops.

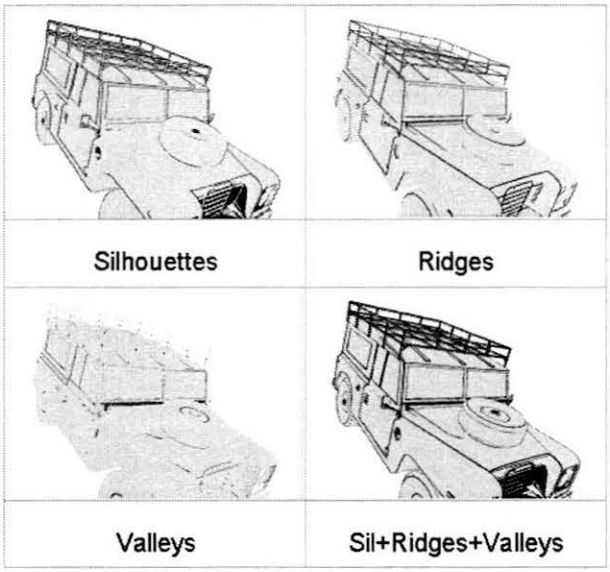
Technical Discussion: We use a camera in the loop to automatically compute the relative projector pose. Our algorithms solve two fundamental problems involved in multi-projector displays. First, we achieve sub-pixel alignment between overlapping projectors by exploiting the homography induced due to the display plane. Second, we use a fast technique to find intensity feathering weights to cross-fade the projected images in the overlapping region. We use 3D graphics hardware to achieve real-time pre-warping of the images so that they generate a single seamless image when projected. The techniques can be used for rear-projector or front-projection display.

Future Direction: We are investigating techniques to achieve color equivalence between overlapping projectors. We also are planning to reduce the calibration time to below 10 seconds.

Contact: Ramesh Raskar, Jeroen van Baar
http://www.merl.com/projects/ProjectorMosaic/

Lab: MERL Cambridge Research
Project Type: Research

Hardware Support for Non-photorealistic Rendering



Special features such as ridges, valleys and silhouettes, of a polygonal scene are usually displayed by explicitly identifying and then rendering 'edges' for the corresponding geometry. The candidate edges are identified using the connectivity information, which requires preprocessing of the data. We present a non-obvious but surprisingly simple to implement technique to render such features without connectivity information or preprocessing. At the hardware level, based only on the vertices of a given flat polygon, we introduce new polygons, with appropriate color, shape and orientation, so that they eventually appear as special features.

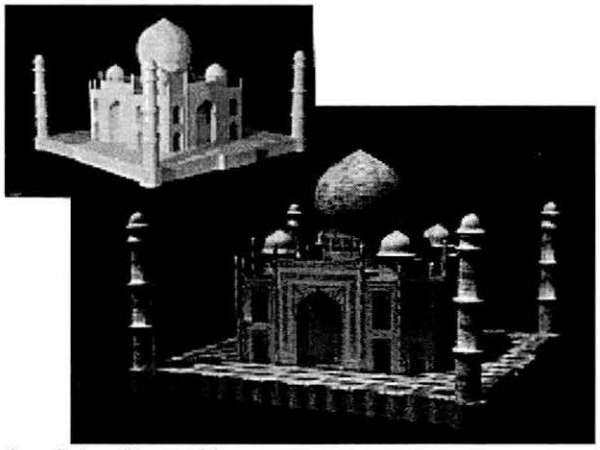
Background and Objectives: Sharp features convey a great deal of information with very few strokes. Technical illustrations, engineering CAD diagrams, cartoons as well as non-photo-realistic rendering techniques exploit these features to enhance the appearance of the underlying graphics models. The most commonly used features are silhouettes, creases and intersections. Most available techniques are restricted to software application implementation. Currently there is no algorithm that can be pipeline or work on small amount of geometric data at a time, preventing hardware support for such features. We focus on an approach that renders special features at uniform width directly into the framebuffer using a primitive shader stage in the graphics pipeline. This allows for the processing of a single polygon at a time, as in traditional rendering, and still display features that would otherwise require connectivity information.

Technical Discussion: We highlight the special features by introducing additional geometric primitives in the scene with no or only small changes in rest of the rendering pipeline, and without using additional rendering operations such as lighting or texture mapping. This type of geometric processing follows the trend of current generation of hardware supporting programmable vertex and fragment operations. The additional polygons are to be generated on-chip, without any software assistance. No data other than the vertex positions and normal of the polygon and current viewing parameters is used. At each frame, we enlarge the back facing polygons for silhouette rendering. Similarly, we add new quadrilaterals at appropriate dihedral angles to all front facing polygons for ridge and valley rendering. The images shown in the figure above are rendered by processing a single polygon at a time, in a software simulator.

Future Direction: We are investigating techniques to extend the support for other effects necessary for cartoons and technical illustrations. We are discussing with designers of 3D rendering chips and graphics APIs for efficient implementation.

Contact: Ramesh Raskar
Lab: MERL Cambridge Research
http://www.merl.com/projects/nonphotorealistic-rendering/
Project Type: Research

Shader Lamps: Animating Real Objects with Projectors



See Color Figure 11

Traditionally, 3D computer graphics is used to modify 3D models using a program and the result is seen on a computer monitor or screen. The typical surface modifications are applying color, changing texture or modifying material properties (bidirectional reflectance) of the object. This invention describes an idea to change the surface appearance of a real object by projecting images. The images themselves are generated using a 3D graphics program and then fed to the projector.

Background and Objectives: Consider a white clay vase sitting on a table. Can we make that vase look like it is made up of gold or marble? Can we change the design pattern on the vase? Can we make the vase to appear as if it is rotating? The perceived color of an object is a function of illumination, surface reflectance and the geometric relationship between the light source(s), surface orientation and viewer location. We can rearrange the factors along the path. For example, a green object illuminated with white light will look the same as white object illuminated with green light. We exploit this idea using a complete 3D geometric representation of the real object and then illuminate it with images with appropriate image intensities.

Technical Discussion: We use a graphics model, which shares the geometry of the physical object but has assigned virtual material properties, for real time rendering. The resultant images are projected on the physical object. Our algorithms solve three fundamental problems involved when a complex object is illuminated by multiple projectors. First, we achieve alignment between the physical object and projected images, semi-automatically using projector calibration techniques. Second, we use 3D graphics hardware to change not just the texture of the object, but also the view dependent appearance. This allows us to make, for example, an object appear shiny with specular highlights. Third, we solve the problem of seamlessly merging images from multiple projectors in the presence of occlusions and self-shadows.

Collaboration: University of North Carolina at Chapel Hill.

Future Direction: We are investigating techniques to extend the paradigm to moving objects and moving projectors. We are looking into applications for illumination of historic architecture, entertainment, laser shows, product showrooms and museums.

Contact: Ramesh Raskar
http://www.merl.com/projects/ShaderLamps/

Lab: MERL Cambridge Research
Project Type: Research

Net Services

Net Services technology leverages the ubiquity and accessability of information provided by the Internet. Internet-ready devices are showing up in automobiles, hospitals, the mobile workplace, and nearly every room in the home. Soon the personal computer will not be the primary means for people to access the Internet. This makes Internet Software a rich area of research and development, as the Internet continues to expand into diverse arenas such as public, private, B2B, B2C, wireless, entertainment, and location aware information services.

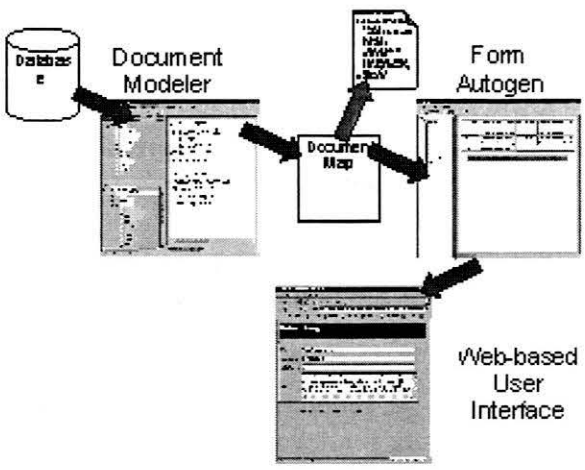
MERL is focused on developing middleware and applications that enable new types of Internet-based services and software products. Our most significant efforts are in the areas on Mobile Agents and Wireless Internet Networking:

- Mobile Agents. MERL as developed a mobile agent technology, called Concordia (page 104), which is currently licensed by several customers. It is a framework for developing and managing network-efficient mobile agent applications useful for accessing information anytime, anywhere and on any device supporting Java. The Concordia system allows custom mobile agents to travel across the Internet, where they can interface with backend applications, databases, or other mobile agents. We are continuing to develop the Concordia framework, making it useful for a wider range of applications, including Enterprise Application Integration.

- Wireless Internet Networking. MERL is developing TWINet, a wireless Internet networking framework that allows for easy, ubiquitous access to personalized information services and content using a personal wireless communication device. TWINet, a technology applied to several MERL projects, is a cell-phone-centric networking technology that turns common consumer appliances, such as a TV set or a car navigation system, into an internet appliance.

Project Descriptions

Concordia XML Framework



Concordia XML is a framework for the rapid development and deployment of Enterprise Application Integration (EAI) applications. The Concordia XML system builds upon Concordia and uses mobile agents to efficiently and flexibly link together enterprise applications and databases. At its core the Concordia XML system includes design-time tools and uses eXtensible Markup Language (XML) to encapsulate enterprise information in a standard form. Concordia mobile agents route XML documents throughout a corporate network based on a customizable set of business rules and access back-end data sources via customized system provided data access adapters.

Background and Objectives: The Concordia XML project, building upon the original Concordia project, has developed a framework which can be used in a variety of business environments to rapidly develop customized EAI solutions.

Technical Discussion: The Concordia XML system is made up of design-time tools for the easy development of customized applications and of a run-time environment for the deployment and execution of these applications. The system builds upon Concordia and is written entirely in Java. The system operates by converting enterprise data from proprietary back-end formats into a platform independent XML format. Concordia agents then route this information around a network, manipulating the information and sharing as required.

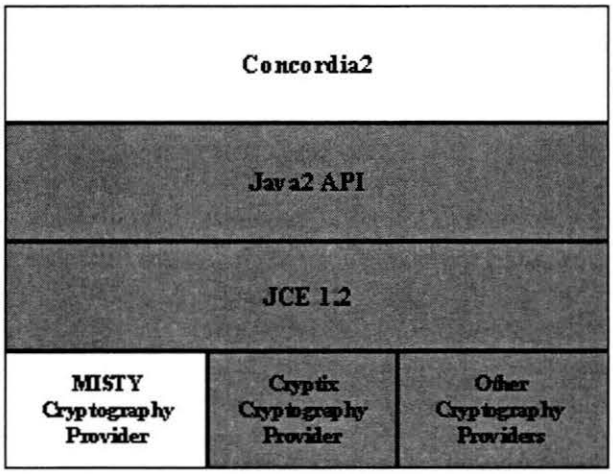
The system's back-end is composed of data access adapters to convert the information contained within legacy data source into XML format. On the front-end, the system uses a web-based user presentation layer that allows end users to view and manipulate information via a web browser. The system's middle tier is composed of a flexible business rules architecture. The system provides design time tools which allow a systems engineer to define how legacy information is mapped into XML format, allows for the automated generation of HTML pages for user presentation, and allows for performance and load simulation, all without requiring knowledge of XML, HTML or even a data source language such as SQL.

Future Direction: Future developments will extend the system by providing generic application templates and additional development tools that can be quickly customized for a specific customer's needs.

Contact: David Wong, Frederick J. Igo, Jr.
http://www.merl.com/projects/concordia-xml/

Lab: MERL Cambridge Systems
Project Type: Advanced Development

Concordia on Java2



The Concordia(tm) Java-based mobile agent systems framework was substantially enhanced to work on the Java2 platform. The key new features in Concordia on Java2 are compliance with the Java2 security model and additional enhancements to the Concordia agent tracking functionality.

Background and Objectives: The Concordia(tm) Java-based mobile agent systems framework is an ongoing technical program at the MERL Cambridge Systems Lab. With the advent of Java2 as the prevalent JDK platform of choice, it became imperative that Concordia follow suit in its support for this particular Java platform.

Technical Discussion: Concordia2 (Concordia on Java2 version) focused on two major areas of improvements: 1) compliance with the Java2 security model and 2) enhanced support for agent tracking.

Security has always been a key strength of Concordia. It has provided a standard of flexible and rigorous access control and data protection far beyond the Java1.x security model. With the release of Java2, many of the standards set by Concordia have now been incorporated into a radically new Java security model and security extensions. In the spirit of maintaining compatibility with mainstream Java, Concordia2 security has been re-implemented to take advantage of the new Java security model and to use the Java Cryptography Extension (JCE) and the Java Authentication and Authorization Service (JAAS). Use of the JCE allows for easy integration of new cryptography resources such as Melco's MISTY.

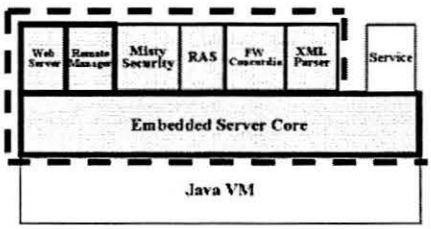
Agent tracking is the functionality in Concordia that allows a user to launch agents in an application and be able track the movements of that agent as it traverses the network on behalf of that user. In prior versions of Concordia, agent tracking required the availability of the RMI stack in the Java runtime in order to be operational. With Concordia2, the user can now track agents using the native TCP support in the JVM. Furthermore, there are substantial improvements in the agent tracking facility to better handle error conditions in Concordia2.

Future Direction: Concordia has grown into a very mature and stable product. Thus, the future direction would focus more on development of new technology in related technology areas which leverages MERL's vast experience gained from this program.

Contact: David Wong, Andrea DeDuck
http://www.merl.com/projects/concordia2/

Lab: MERL Cambridge Systems
Project Type: Advanced Development

Concordia for Power Systems Business Units



A major success story for the Concordia mobile agent program has been its adoption by MELCO's Power Systems Business Units (Sanden and Keito-bu) as a core technology in their solutions for the electric power industry. For the past few years, much of the Concordia development has focused on meeting their requirements and on effectively transferring this technology to MELCO for formal product customization, maintenance, and support in Japan.

Background and Objectives: The Concordia mobile agent technology has been well received by MELCO's Power Systems Business Units, Sanden and Keito-bu, and has become a core technology in their systems solutions for the electric power industry. In the past few years, the Concordia program has focused much of its development efforts specifically to meet requirements from these two business units. Furthermore, we have conducted thorough technology transfer to Sanden's affiliate company, MCR, in order that formal product customization, maintenance, and support for Concordia can be established in Japan.

Technical Discussion: Some of the important Concordia product features that have been developed specifically for Sanden and Keito-bu include: 1) streamlining of the Concordia core, 2) upgrading Concordia to Java2, 3) enhancing Concordia for wide area network support, and 4) transferring the Concordia source code and requisite technology know-how to Sanden and MCR.

The effort on streamlining the Concordia core and reducing its memory footprint resulted in the development of Lightweight Concordia, which was then transferred to Sanden with extensive training from the Concordia team. Subsequently, Sanden further reduced the memory footprint of this Lightweight Concordia version into a Featherweight version of Concordia for embedded home metering devices.

The development of Concordia on the Java2 (Concordia2) platform was also primarily driven by requirements from Sanden. A separate description of this version of Concordia can be found elsewhere in this report.

The work required to make Concordia more suitable to wide area networks and the Internet included improving the error handling capabilities of the Concordia runtime and making it much easier for enterprise users to transmit Concordia agents through existing corporate firewalls.

Finally, substantial effort was expended by both MERL and Sanden to ensure that the Concordia technology was transferred effectively to MELCO. Many Sanden and MCR engineers visited MERL for up to 6 staff-months each to work with the Concordia team in order to acquire the appropriate know-how to maintain and support the Concordia product in Japan.

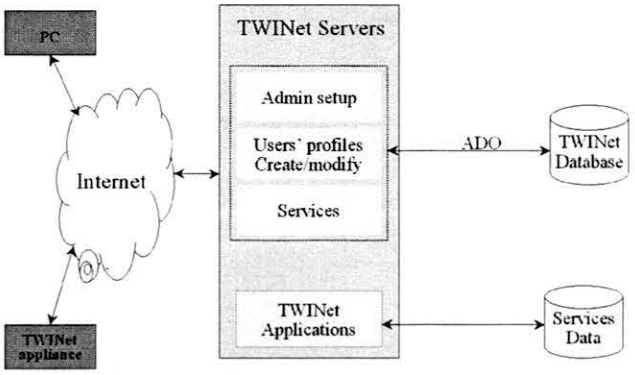
Collaboration: The Concordia program has been working very closely with Sanden for the past three years.

Future Direction: We will continue to serve in the role of being the primary mobile agent technology provider to MELCO's business units and, in particular, Sanden and Keito-bu.

Contact: David Wong **Lab:** MERL Cambridge Systems
http://www.merl.com/projects/concordiaSanden/ **Project Type:** Advanced Development

Internet Server for TWINet Application

TWINet Server Architecture



TWINet is a wireless networking framework that allows for easy, ubiquitous access to personalized information services and content using a personal wireless communication device, using any compatible appliance such as a TV set or a car navigation system for the rendering. The users access their personalized services such as weather condition, stock prices, picture album, etc. at home on a TV set, or in the car through the sound system, using their cellular phones plugged into connector devices. The services are configured for each user and for each client device. The server side stores personal profiles and services, manages content brokerage, transmission and storage, and interacts with the clients.

Background and Objectives: The TWINet Server (TWINServer) constitutes the server side of the TWINet framework. It manages the user and device profiles, information gathering and storage for the services, and the transmission to the clients. The objective is to provide to the users high value services that they can access anywhere on any compatible appliance.

Technical Discussion: The TWINet server side comprises of databases that store user and services information, applications that provide services, and interfaces for the administration of the user profiles and the overall system. The applications gather traffic, stocks, etc. data from the Internet based on user requirements. The e-mail messages are also managed by the server side applications. Currently supported services include weather, traffic, stocks, lottery, e-mail, to-do list, photo gallery, etc. Available services and device types are stored in the database. The users' database store personal information, preferences for each selected device and for each selected service. Currently, we target the TV and the car navigation system as the primary client devices.

Collaboration: IESL (Sanken), MEAA (Mitsubishi Electric Automotive America)

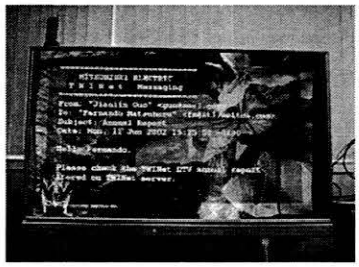
Future Direction: Improving the services provided by TWINet and adding new ones is the main course for improvement. New issues will need to be addressed as new multimedia services are added. In the future, information and content can be directly obtained from content providers through contracts. Transmission of content via alternative paths such as cable, digital broadcast, as well as wireless Internet is also in the scope of the TWINet architecture.

The server will need to support a very large number of users in the future. We may need to work with a web portal, or continue with a small version of the server. We may need to consider load balancing and similar issues. We will also improve the administrative aspects of the server.

Contact: Johnas Cukier
http://www.merl.com/projects/TWINet-server/

Lab: MERL Murray Hill
Project Type: Advanced Development

TWINet Digital Television



DTV brings to TV much more capabilities than ever before. One of the most attractive ones is the Internet. However, the question is how to easily turn DTV into an Internet device without any complicated wires and devices. TWINet DTV is a viable answer. The only device needed is a cellphone, which will be carried any time, anywhere. TWINet DTV supports intelligent programming selection, Internet navigation, and personalized information simultaneously via the same high quality TV screen. No wires, no PC, no more small display of mobile device, no extra skills needed to set up. Don't waste time to watch less interesting TV advertisements, only information you are interested in.

Background and Objectives: TWINet DTV is an application of the TWINet. The objective is to provide an easy-to-use Internet solution by simply utilizing the DTV and the cellular phone. The interconnection is as easy as plugging the cellular phone into a DTV docking station. DTV turns into an Internet device. The phone is charged while plugged-in. When un-plugged, the cellular phone and the DTV can be operated independently. The TWINet DTV has market potential to increase sales of the DTV and cellular phone, build partnership with DTV and Internet service providers.

Technical Discussion: To provide an easy no extra cost Internet access solution, utilize the high quality DTV display and overcome the small display size of cellular phone, we have proposed TWINet DTV project to leverage MELCO strengths in DTV and cellular technologies to the Wireless Internet Domain. Our strategy is to enhance existing infrastructures and launch quick product. A digital communication protocol TWINLink between DTV and cellular phone has been implemented by utilizing the serial technologies and a TWINdock has been designed to physically connect DTV and the phone. The cellular upstream channel is used for sending data to service provider. For downstream data, the cellular downstream channel or other available fast channel (cable, satellite, etc.) may be adapted. The user information is associated with the cellular phone and therefore no explicit dialogue is required for getting the TWINet DTV services. A common user profile (preferences, geographical region, etc.) is available to all services. With TWINet DTV, TV program and TWINet data (message, stock, family picture, web pages, etc.) can be displayed simultaneously. TWINet data can be displayed with transparent background (TV video is visible) or opaque background (TV video is blocked). The different display font and color can also be chosen. User will be notified automatically if there is a message. User may view TWINet data at any time by just pressing TV remote control buttons.

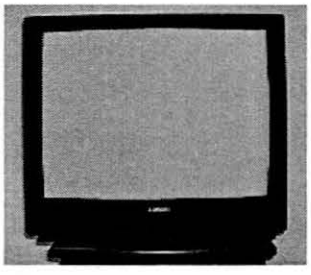
Collaboration: Mitsubishi Electric Digital Equipment America, Inc.

Future Direction: Our current effort is to improve the quality of services by compromising the unreliability of wireless connection, enhance the TWINet DTV functionality and work on the web browser. The real time digital broadcasting will be our next objective. We will also explore the use of Bluetooth and eventually wireless 1394 for the future generations of TWINet DTV.

Contact: Jainlin Guo
http://www.merl.com/projects/TWINet-DTV/

Lab: MERL Murray Hill
Project Type: Advanced Development

Wireless Internet Connection to Legacy Televisions



One of the TWINet project's goals is to provide simple to use wireless Internet access with the television as the primary display. While new digital televisions can be designed to provide TWINet functions out of the box, legacy television sets having only analog signal inputs will need an adapter in the form of a set-top-box to perform the tasks of connecting to the mobile phone, rendering the graphic and audio presentation on the TV, and accepting user inputs.

MHL will design and produce a prototype set-top-box demonstrating the TWINet concept implemented on a legacy television using only analog video and audio connections, while keeping the mass production cost of this device low to trigger the proliferation of these terminals.

Background and Objectives: The TWINet concept aims to provide easy to use personalized Internet access through one's digital mobile phone, spanning across many environments a person may encounter such as inside a vehicle, hotel room, airport, relative/friend's house, or one's own living room. Most of these environments will initially have a present day analog television to serve as the display. Therefore it is paramount to have an inexpensive adapter to enable TWINet services on these televisions to spearhead a speedy growth of users.

Technical Discussion: An adapter providing TWINet functions on a legacy television must be capable of connecting to any data-enabled mobile phone, identifying the phone's owner, supporting digital communication using Internet Protocol, and rendering graphical user interface constructed on an Internet browser. User input will include a simplified hand held remote control with the option of a full size computer keyboard. The connection to the television will utilize only its analog video and audio inputs. Additionally, further research will be made on whether to include other value-added features on this set-top-box in order to provide incentives for the purchase and utilization of this terminal.

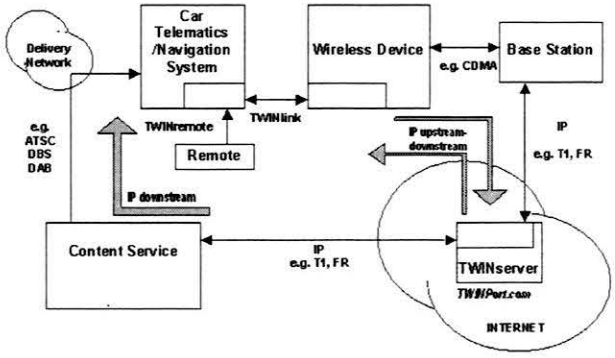
Collaboration: MELCO-Industrial Systems Lab.

Future Direction: MHL will try to leverage expertise from several MELCO sources on the development and adaptation of a Windows CE platform capable of achieving the goals set forth above and produce a demonstration by October 2001.

Contact: George Fang
http://www.merl.com/projects/TWINet-analogTV/

Lab: MERL Murray Hill
Project Type: Advanced Development

Wireless Connections to Car Navigation Systems



With the explosive growth of the Internet for both business and personal needs, easy-to-use connectivity solutions are required. TWINet can be used to gain access from various locations without the need to involve specialized equipment such as a PC, complicated wiring, or limitations of a small display in current mobile devices. TWINet easily turns any appliance such as TVs and car telematics systems into Internet connected devices by using a cellular phone.

Background and Objectives: This project concentrates on the car telematics/navigation systems used with TWINet.

Technical Discussion: The key feature of Wireless Connections to Car Navigation is the ability for the user to bring his personal network environment to the car. Here the user is quickly brought to a familiar environment that spans across the office, home, and home away from home environment. The user also has access to pertinent data relative to the automobile experience. In other words, the TWINet system is sensitive to the user's environment as well. The system identifies four basic entities: the user, the appliance (e.g. car telematics, TV, etc.), the available communications channels, and the services available to the user. For the automobile, the typical components available in the car might be a: car navigation system, global positioning system, and various car sensors (engine diagnostics, speed, tire sensors, fuel level indicator, etc.). The availability of other components that are not typical today would include local storage for services such as multimedia files (movies, games, maps, guidebooks, etc.).

The automobile interface is another issue. The main concern for the driver is that he is not distracted while using the system. For that reason, voice interaction may play a key role for the automobile interface. User interface components include: text-to-speech, voice command recognition and a simple one to three button interface that is easily accessible by the driver. The passenger interface may include a richer set of components such as: high-resolution screens/touchscreens, keyboards, mice, game controllers, infrared remote controller, etc.

Collaboration: IESL (Sanken), MEAA (Mitsubishi Electric Automotive America).

Future Direction: In the future, wireless wide-area (Digital Audio Broadcasting) and narrow-area (IEEE 802.11 standard) broadcasting will be available to the motoring public. The automobile system should take advantage of both developments for streaming multimedia to the car.

Contact: Johnas Cukier
http://www.merl.com/projects/TWINet-car-nav/

Lab: MERL Murray Hill
Project Type: Advanced Development

Web Page Services

The Web Page Services project is a lightweight framework for providing and applying services or capabilities to the web page that is currently loaded into a web browser. In this way, after a user browses to a web page, the usefulness of that web page can be dynamically enhanced. For example, a particular Web Page Service might allow a browser user to create a chart from a column of prices in a table on that web page. Another service might allow the user to manipulate (resize or replace) an image in that web page. Yet another service might insert real-time stock quotes wherever a company name occurs in the page. Since the implementation is lightweight, the client web browser does not require any Java support, browser plug-ins, or other programs to be pre-installed.

Background and Objectives: This effort improves the usefulness of existing web pages. Rather than incorporating sophisticated Dynamic HTML capabilities into each page on the web site, this framework allows such capabilities to be dynamically inserted into a web page as needed after the user has navigated to a page. Unlike similar browser tools, this system modifies the contents of the current page rather than re-directing the user to a new page. This provides for a more seamless and satisfactory browsing experience since users can dynamically tailor the content they view to their individual tastes or requirements. Furthermore, this effort does not require a plug-in or a program to be installed, which facilitates usability and convenience -- and greatly alleviates user security concerns.

Technical Discussion: This effort uses advanced scripting techniques in a companion browser window or frame to dynamically analyze the content of the currently-loaded web page. That web page can then be enhanced by leveraging one of three sources of information: 1> Advanced abilities built into the web browser itself (such as the ability to find and highlight text strings in the page); 2> Existing web services (such as external stock quote or currency conversion services); 3> Back-end hardware or software resources (such as inserting a link from an employee's name in the web page to information about that employee in a back-end corporate employee database). This project also leverages forthcoming (initially SOAP-based) web services architectures from numerous software vendors (Microsoft .NET, Sun ONE, etc.).

Collaboration: Collaboration is anticipated with business units that have web-based content that can benefit from domain-specific services.

Future Direction: Explore alternative environments and domain-specific web page services.

Contact: Alan Esenther
http://www.merl.com/projects/wps/

Lab: MERL Cambridge Systems
Project Type: Initial Investigation

Networks

As computers have become smaller and more powerful, almost all devices have computing power in some form. Therefore, organizing a collection of such devices – at home, work, or mobile environments – requires development of networking technologies that enable command and control of a collection of devices, not all of which are computers. Such technologies have to go beyond traditional computer networking techniques.

In the area of Networks, MERL is exploring several different types of technologies for consumer and industrial uses:

- UPnP: MERL is an active participant, contributor, and proponent of Universal Plug and Play (UPnP), an open internet architecture for peer-to-peer network connectivity of intelligent appliances, wireless devices and PCs of all form factors.

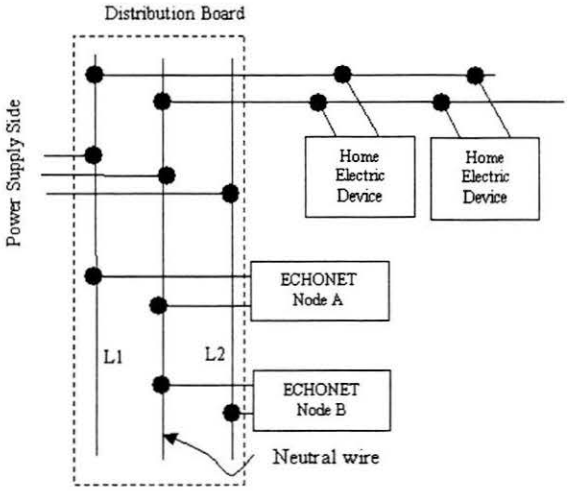
- PLC: We are developing Power Line Communication (PLC) technology, which seeks to transmit a few kilobits to a megabit per second of data over normal AC power lines. The idea is to make use of transmission bandwidth potential available in the existing power line wire infrastructure.

- LVDS: Low Voltage Differential Signaling (LVDS) is a high speed I/O technology for communication between silicon chips. We are currently monitoring developments in this high speed technology, with an eye toward applying them to improve Mitsubishi Electric products and services.

- DTV: We work on both the broadcast and the consumer sides of networking Digital Television (DTV). On the broadcasting side, we support the High-Definition encoders made by Mitsubishi Electric. On the consumer side, we work with DTV manufacturing groups in the US on integration of networking technologies (such as IEEE 1394, HAVi, etc.) into the DTV framework.

Project Descriptions

PLC LSI Development



Power Line Topology
(*Single-phase 3-wire type*)

Power Line Communication (PLC) physical layer LSI is based on ECHONET specification - the modulation type is Discrete Multi Tone (DMT) and transmit rate is 24kbps. PLC DSP and FPGA are currently using floating-point calculation. Fixed-point solution is one of the ideas to reduce the cost. The other solution could be using DWT/IDWT pairs in modulators and demodulators rather than IFFT/FFT pairs to recover the original transmitted signal with lower BER. The project involves examining currently available floating-point implementation and it's performance, and introducing methods both to reduce the overhead of the current implementation and to gain more robustness of the power line communications. The project is indispensable for lowering the cost of PLC physical layer LSI.

Background and Objectives: ECHONET was created to realize in-home communications infrastructure providing fast, high-bandwidth transmission of data and images, at the same time incorporating a relatively low-speed/bandwidth/cost network that is compatible with conventional home appliances and equipment. This network will enable the interconnection and systematic operation of a wide assortment of home appliances and controllers from different manufacturers. It aims data transmission without special wiring, easy development of multi-vendor home systems, response to long lifetime and home system proliferation, connectability and coexistence with other (AVC) systems.

Technical Discussion: The ECHONET supports PLC-B system that performs communications by putting modulated signals on multiple tones spread at equal frequency spacing of 4.3125kHz. The tone frequency is caused to hop in an adaptable form depending on the transmission line condition. Primary modulation system used in DMT is one of DQPSK/DBPSK/D8PSK. It is proven that in OFDM technology that DWMT performs better than DMT. Therefore, we expect DWMT mod/demod designs might also improve the performance of PLC communications.

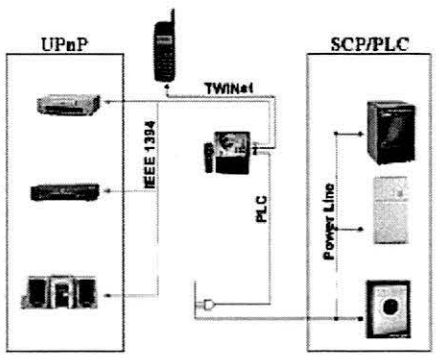
Collaboration: MELCO L-ji-se.

Future Direction: Initially, analysis and evaluation of source code and the floating-point implementation. Afterwards, a fixed-point model will be implemented in Matlab environment. Future R&D directions include the quantification of the performance improvement of the transceiver unit after source code modification.

Contact: Zafer Sahinoglu, Johnas Cukier
http://www.merl.com/projects/PLC/

Lab: MERL Murray Hill
Project Type: Advanced Development

Simple Control Protocol (SCP)



Home Multi-network Connectivity: UPnP, SCP & TWINet

Simple Control Protocol (SCP) is a lightweight networking protocol, proposed by Microsoft Corporation, which allows the manufacturers to produce small and intelligent devices. It will use Power Line Carrier (PLC) as its physical layer. SCP defines the terms of physical and logical networks. In a building complex, for example, the physical network consists of all devices connected to the power line. The devices in each apartment make a logical network. The system provides mechanisms for defining and/or deleting logical networks and providing adequate security. The network, however, has a very limited bandwidth, i.e. about 4 kbps, that allows limited capability for data communication. An SCP network is shown on the right half of the diagram above.

Background and Objectives: Home devices fall into two classes based on the their price, processing power and data requirements. The first category consists of electronic home devices such as a TV, Jukebox, or DVD player, which may generate a large amount of data across the network. On the other hand, electric appliances generate a little amount of network data. The main objective of networking these devices is to provide automated operational control, diagnostic, and remote updating of firmware, etc. While the TCP/IP based UPnP initiative is the main thrust for electronic devices, SCP/PLC is currently being seen as the networking solution for electrical appliances.

The objective of this project is to get SCP and PLC technologies and Mitsubishi M16C processor integrated on a single chip. That chip can be supplied to device manufacturers at a low cost.

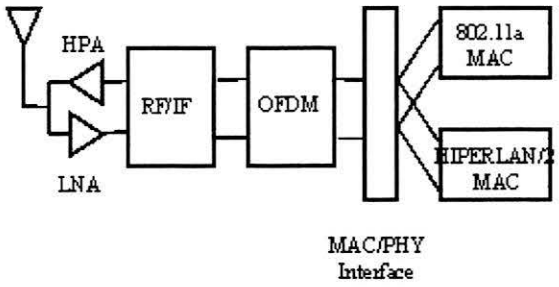
Technical Discussion: SCP aims at providing an inexpensive solution due to its simple architecture, small memory footprint, and using the existing infrastructure (electric wiring). But it has placed some severe restrictions on its scope due to these very reasons. The limited bandwidth of physical layer (i.e. PLC) is the most serious one. Besides, the SCP architecture is a non-traditional one. It has a dual stack which makes it complex. There is a special entity, called the Address Space Arbitrator (ASA), which creates logical networks. An SCP network does not have direct access to the Internet. Rather it uses a bridge to UPnP, which can access the Internet.

Collaboration: MHL is working closely with MEUS DEC-E and MELCO Han-pon (L-ji-se) on the SCP project.

Future Direction: Home networking offers a big opportunity for manufacturers of intelligent devices. The SCP chip solution can be supplied the manufacturers of appliances. A SCP-UPnP bridge can be implemented Mitsubishi devices.

Contact: Fernando Matsubara, Ghulam Bhatti, Johnas Cukier **Lab:** MERL Murray Hill
http://www.merl.com/projects/SCP/ **Project Type:** Advanced Development

An Universal MAC/PHY Interface for Different WLAN Standards



Wireless LAN (WLAN) is widely considered to play a major role in wireless multimedia communications. Two competing standards: IEEE802.11a and HIPERLAN/2 will dominate WLAN market. These two standards have a similar physical layer that is based on Orthogonal Frequency Division Multiplexing (OFDM) technology, but the Medium Access (MAC) layers are different. MELCO is currently developing the chipset for WLAN. MELCO has strong capabilities in the radio front end as well as OFDM implementations. The project here is to develop a universal interface between MAC and Physical (PHY) layers. This allows the reuse more of PHY components in both IEEE802.11a and HIPERLAN/2 products, and helps MELCO's position in WLAN chip market. In this way several WLAN manufactures could use the MELCO PHY chipset and connect any of the WLAN MAC layers to this PHY chipset.

Background and Objectives: Two different incompatible WLAN standards operating in the same 5GHz band - IEEE802.11a and HIPERLAN/2 will cause interferences and other problems. From the end user perspective, the end user is forced to make a standard choice, and no roaming is possible. From the manufacture perspective, two sets of different chipsets have to be made for the WLAN market. Now, IEEE 802.11 working group is developing algorithms that allow inter-working between these two standards. The manufactures are trying to develop a single universal device that is IEEE802.11a and HIPERLAN/2 compliant.

Technical Discussion: Since IEEE802.11a and HIPERLAN/2 have quite similar PHY layers and much different MAC layers. The goal of the universal MAC/PHY interface is to try to reuse more PHY components in both 802.11a and HIPERLAN/2 products. The basic idea is that the MAC/PHY interface will include two training sequence generators. One is for both 802.11a and HIPERLAN/2; the other is a short training sequence generator for only HIPERLAN/2. The interface will also contains an initialization procedure that can initialize both 802.11a MAC data and HIPERLAN/2 MAC data unit for physical transmission. In addition, the convolution en/decoder in PHY layer has to be designed to offer the code rates of 1/2, 9/16, 2/3,3/4, and extra puncture scheme for HIPERLAN/2.

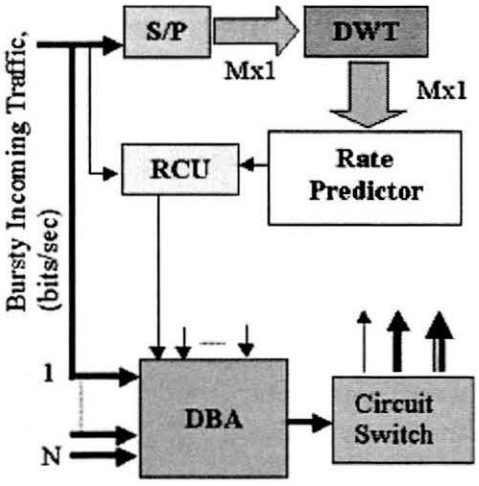
Collaboration: MERL-MHL is actively collaborating with L-ji-se System Development Group.

Future Direction: Future work will focus on refining and modifying the MAC/PHY interface, and developing the algorithms for data rate adaptation.

Contact: Daqing Gu, Philip Orlik
http://www.merl.com/projects/wireless-LAN/

Lab: MERL Murray Hill
Project Type: Advanced Development

Network Traffic Management



The provision of QoS to applications traffic depends heavily on how different traffic types are categorized and classified, and how the prioritization of these applications are managed. Bandwidth is the most scarce network resource, and therefore there is a need for a method or system that shares the available bandwidth in a network among different applications in such a way that each class or type of traffic receives its constraint QoS requirements.

This project consists of three autonomous sections: the design of traffic bit arrival rate predictor, the development of optimum bandwidth renegotiator by proper design of RCU, and the design of circuit selector to navigate a service admission request among different transmission media.

Abbreviations: DWT: Discrete Wavelet Transform, RCU: Renegotiation Control Unit.

Background and Objectives: In this project, dynamic resource allocation techniques to VBR bursty traffic are introduced. The DBA (Dynamic Bandwidth Allocation), upon receiving a renegotiation flag from RCU, assigns available bandwidth on one of the transmission media to the newly requested service. The proposed predictor and the bandwidth renegotiation methods are online and practically embeddable to QoS management blocks, routers in networks and Digital Subscriber Line Access Multiplexers (DSLAM) in DSL access systems. Also, we aim to use the gained traffic management and service admission control techniques for TWINET project.

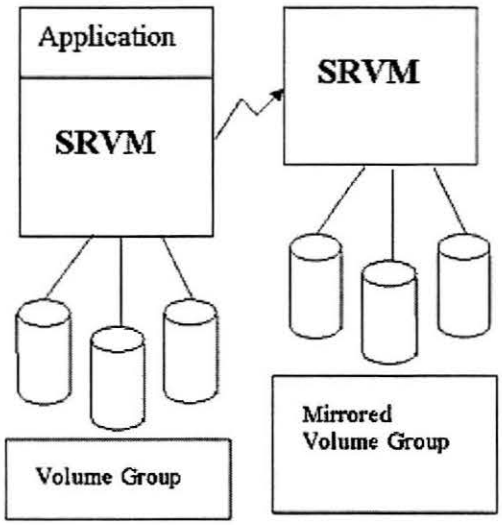
Technical Discussion: The empirical results show that on-line dynamic bandwidth allocation to applications traffic by the introduced predictor outperforms traditional dynamic bandwidth allocation methods and adaptive algorithms such as LMS and RLS in terms of utilization and queuing delay. The renegotiation frequency is a trade-off between signaling overhead and high bandwidth utilization. High renegotiation frequency loads the network with heavy overhead. On the other hand, long inter-renegotiation intervals make the follow-up of the traffic bit rate pattern difficult. RCU determines the optimum number of bandwidth renegotiations given buffer size, renegotiation cost, and bandwidth costs as inputs.

Future Direction: The future work will be the adaptation of the developed dynamic bandwidth allocation and dynamic circuit assignment techniques to the TWINET project in traffic management to maintain the applications required QoS.

Contact: Zafer Sahinoglu
http://www.merl.com/projects/TrafficMgmt/

Lab: MERL Murray Hill
Project Type: Advanced Development

Network Replication



The Network Replication project is a cooperative effort with Veritas Software to provide network replication capability to their popular Volume Manager product, VxVM. Our technology has been licensed by Veritas for their Veritas Volume Replicator product (VVR).

The replication engine of VVR offers an ideal solution for the remote archiving of storage such as databases or file systems through its ability to replicate a virtually unlimited number of related data volumes while maintaining consistency. Replication uses standard networks without proprietary hardware and is highly resistant to system and network failure.

Veritas Software is a market leader in storage management software.

Background and Objectives: The Network Replication project evolved from an R&D exploration of network distributed storage. Establishment of our relationship with Veritas Software refined this objective to network replication of Veritas' logical volumes.

Technical Discussion: Veritas' customers use VxVM to protect their data from media failure by creating local mirrors, but remote mirrors were needed to protect against system and infrastructure failure. VVR replicates groups of logical volumes, maintaining consistency among them during replication. Changes to any member of the volume group are transparently captured and replicated to one or more remote locations.

VVR has flexible replication and configuration characteristics. It can replicate synchronously or asynchronously. Feedback allows input flow rates to be throttled if necessary to match available network bandwidth. It can simultaneously replicate volume groups to remote sites while acting as a receptacle for volume groups replicated from other sites and it can support an unlimited number of volume groups. It can replicate a volume group to multiple destinations, each with independent replication characteristics and latencies. Recent work has focused on improving the ease and speed of the failover/failback procedures.

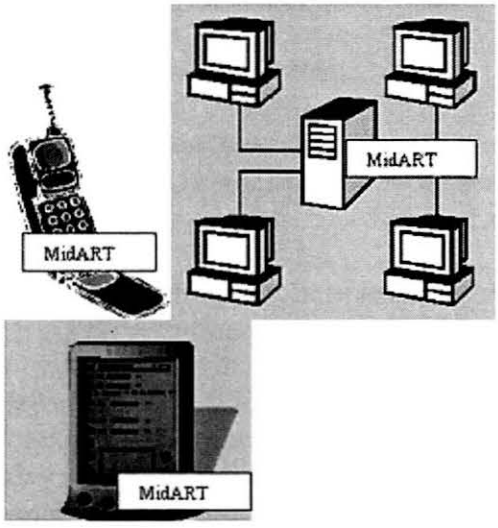
Collaboration: Since 1997, CSL has worked exclusively with Veritas Software in the development of VVR, deriving market direction and market access from this relationship.

Future Direction: The base VVR product will be ported to new platforms such as Windows NT and AIX, and its feature set expanded to better serve the disaster recovery market.

Contact: David Rudolph
http://www.merl.com/projects/netrep/

Lab: MERL Cambridge Systems
Project Type: Advanced Development

Real-Time Middleware for Heterogeneous Network Environment



MidART is a distributed real-time middleware software package with easy-to-use programming interface for multimedia, data acquisition and message communication over UDP/IP. It allows rapid development of multi-threaded concurrent communicating applications in a network environment with Quality of Service provision. MidART Version 1.0(NT) for Windows NT 4.0 has been released. So far, eight universities have signed MidART source license agreement to use MidART in various research projects. The communication services developed in MidART have also been used in Sanden's industrial plant control products. The existing MidART is the result of a collaboration with Sanken and the University of Massachusetts, Amherst.

Background and Objectives: Existing network software facilities such as the socket interface is cumbersome and difficult to use for application builders. Moreover, real-time applications need end-to-end quality of service provision. To facilitate the construction of distributed real-time applications on open off-the-shelf systems, there is a strong need to first provide easy-to-use real-time programming models and services to real-time application designers. The existing MidART fills this need for homogeneous network systems. Our ongoing work is to enhance MidART in order to support applications on heterogeneous networks that include desktops, as well as handheld and wireless devices.

Technical Discussion: MidART provides a set of real-time application specific but network transparent programming abstractions that support individual application's data acquisition, communication and QoS requirements. The focus of the middleware is to support the end-to-end application real-time data transfer requirements with a set of easy-to-use communication service programming interfaces. The key services provided by MidART are Real-Time Channel-based Reflective Memory (RT-CRM) and Selective Channels. RT-CRM is a software-based reflective memory similar to the producer/consumer model. Selective Channels allow applications to dynamically choose the remote node(s) which data is to be viewed from and sent to at run time via a set of channel start and stop protocols.

Collaboration: The current MidART research is in collaboration with Prof. Gerhard Fohler at the Department of Computer Engineering, Malardalen University, Sweden.

Future Direction: MidART is being ported onto Windows CE for handheld and mobile devices over wireless Ethernet. We are experimenting with video streaming applications, developing methods for heterogeneous system where non-real-time nodes do not need to install MidART, and developing a plug-in based modular scheduler for host and network scheduling.

Contact: Chia Shen
http://www.merl.com/projects/midart/

Lab: MERL Cambridge Research
Project Type: Advanced Development

Spoken Language Interfaces

Speech recognition and speech synthesis technologies have been available in crude form for nearly two decades. However, due to technical limitations, their application has been limited to a few success stories. Meanwhile, over the same two decades, a revolution in consumer electronics and computing devices has dramatically increased the market need for Spoken Language Interfaces in order to simplify UI and free-up the hands and eyes. So while speech recognition and synthesis technologies continue to make incremental improvements, the potential for a revolution in Spoken Language Interfaces looks promising.

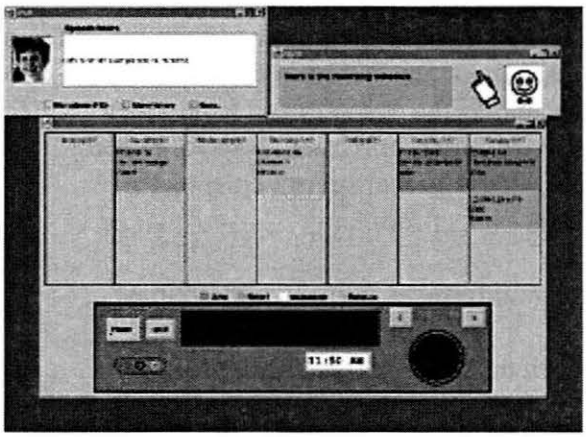
At MERL, our approach to research in Spoken Language Interfaces takes two directions:

- Speech-Centric Devices. Most applications of speech technology take the approach of layering a speech interface on top of existing hardware, providing the user the option of pressing buttons with either the finger or the voice. At MERL, we are interested in creating new devices designed with Spoken Language Interfaces in mind. We believe that the result will be improvements in industrial design, hardware costs, and interface accuracy.

- Conversational Speech Interfaces. The poor accuracy of even the best speech recognition programs has made it impossible to consider using speech as the primary interface for complex tasks. Ironically, it is the complex task that needs a speech interface the most. At MERL, we are applying principles from human collaborative discourse theory to build a conversational framework on top of which we can layer speech interfaces. With this framework, we are able to build into the software system some understanding of the task at hand. We believe that this will make conversational, natural Spoke Language Interfaces more accurate, robust, and useful for real products.

Project Descriptions

Collaborative Spoken-Language Interfaces



We have developed a prototype spoken-language interface to programming a video-cassette recorder (VCR), in which a user collaborates with an intelligent software agent in natural spoken English. In the collaborative conversation, the user communicates his goals to the software agent and the software agent asks for more information when necessary. The agent takes care of the details of managing the recording schedule and operating the VCR controls. This prototype demonstrates generic technology we have developed for building collaborative spoken-language interfaces for varied applications.

Background and Objectives: Speech recognition technology has matured dramatically in the past few years, with the first generation of products with embedded speech recognition now coming to market. These products typically use a small set of command words as an alternative to pushing the buttons on a cell phone, car dashboard, or other device. Our goal is a second generation of spoken-language interfaces, which will support much more complex interactions, such as programming a VCR, operating a home network, or retrieving information from a large database.

Technical Discussion: The architecture of a collaborative spoken-language interface has three main components: spoken-language understanding (which starts with speech recognition and then maps from sequences of words to their meanings), spoken-language generation (which maps from meaning representations to sequences of words and then to speech), and collaboration management (which maintains a model of the conversation structure and the task status). We are currently using commercially available products for the first two components and the COLLAGEN (COLLaborative AGENt) system for the third.

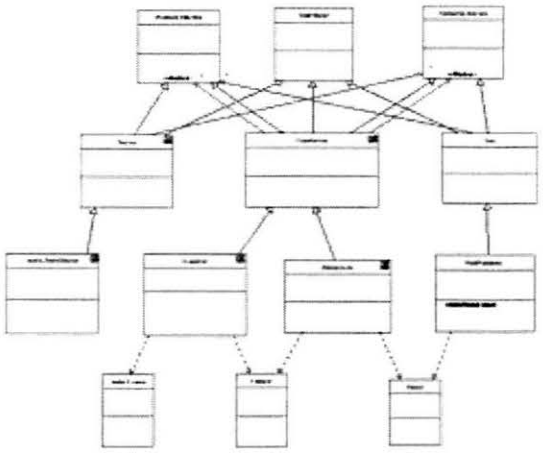
Each of the three components above operates by applying application-independent algorithms, such as parsing or plan recognition, to application- and/or language-specific data sets, such as a grammar of English or a hierarchical task model for VCR programming. Assembling complete data sets for all of these components for a particular application is a very large engineering effort, which we have not undertaken for the VCR programming prototype.

Future Direction: We plan to improve the robustness of the VCR programming demonstration enough to support preliminary user studies, as well as to investigate opportunities for other applications. In the basic research direction, we are planning to investigate ways to use the discourse context (provided by COLLAGEN) to reduce the speech recognition error rate and to support more natural error-recovery subdialogues.

Contact: Candace Sidner, Charles Rich
http://www.merl.com/projects/csli/

Lab: MERL Cambridge Research
Project Type: Research

SPIEL Toolkit



The SPIEL Toolkit is middleware for Speech applications. It enables the rapid construction of new applications and provides a common API that separates applications from the specific APIs of the speech engines. This also permits fair performance comparisons of speech recognition engines. Specified in UML, Spiel is platform and programming language independent, and especially suitable for embedded, server, and distributed applications.

Background and Objectives: Developing code that controls the speech recognition process is costly, as it must concurrently control the capture of audio data, signal processing, pattern matching, and natural language processing in a limited memory and processor footprint. Further, the probabilistic behavior of speech recognition engines requires special logging support for reproducible testing and tuning of applications.

Much of this logic is common between applications. Our objective is to identify and abstract this shared behavior as a common architecture. This greatly simplifies the production of robust and efficient applications. Since applications are not tied to a particular engine, this also enables a "best of breed" approach in choosing speech engines.

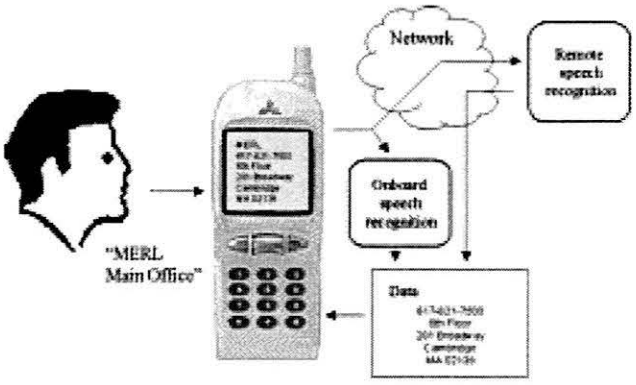
Technical Discussion: For a number of reasons, current industry standard speech middleware, such as SAPI from Microsoft and JSAPI from Sun, are not suitable for embedded, distributed or server applications. Unlike SAPI and JSAPI, SPIEL does not tie the application to an OS or language platform. Since SPIEL is specified in UML, it can be produced as C, C++ or Java, and on Windows, Linux, WINCE, and many embedded platforms.

Future Direction: In order to make SPIEL more useful for prototyping and evaluating speech applications and engines, we plan to extend it by implementing interfaces for additional speech recognition, compression and text to speech engines.

Contact: Peter Wolf
http://www.merl.com/projects/SPIEL/

Lab: MERL Cambridge Systems
Project Type: Advanced Development

Speech Enabled Cellphone



Speech Recognition and our understanding of Speech User Interfaces have matured in the last few years. The Speech Enabled Cellphone project is a targeted effort to create improved cellphone user interfaces and create new products and services by incorporating speech technologies. The possibilities range from enhancing simple interfaces like those in current cellphones, where speech adds convenience, to accessing distant data and media via a handheld device, to creating completely new speech-centric devices where speech is the only input mechanism.

Background and Objectives: Recently, very basic speech technology (voice name dialing) has begun to appear in cellphones as a differentiating feature. We believe it is a successful feature, and that it will become more popular. Therefore, the objective of the Speech Enabled Cellphone project is threefold: to predict speech feature trends in the cellphone market, to develop new speech features and interfaces which can be used in MELCO products, and to evaluate the costs and benefits of such features.

Technical Discussion: We expect processing power resident on cellphones to rapidly increase over time, and new distributed services will also become possible. This will enable a large range of new speech applications on the cellphone platform. We plan to prototype and evaluate some examples of these new products and services.

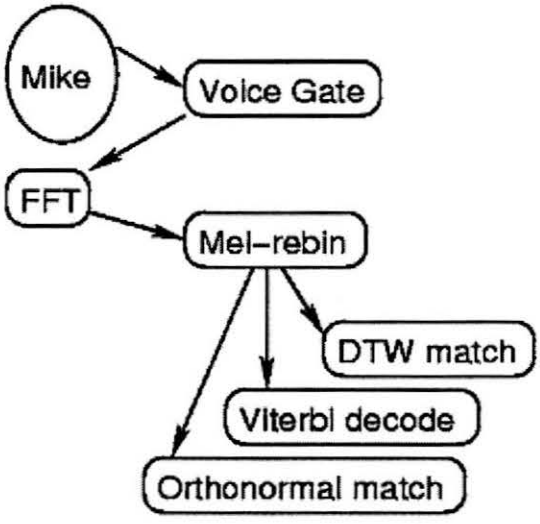
However, existing cellphones have enough computational resources to support onboard small-vocabulary speech recognition systems. We are also developing an onboard system that locally supports connected digit dialing and command and control functionality as an extension of current button interfaces. We will extend this system to build a speech-centric cellphone which relies solely on speech as the interface and does away with the need for buttons and a display. In the longer term, we will also incorporate distributed speech recognition to build systems capable of database query and voice messaging.

Future Direction: Future systems could incorporate noise reduction technology and onboard language modeling in the handset. We will also explore distributed recognition systems and the capabilities made available by large-vocabulary systems.

Contact: Peter Wolf
http://www.merl.com/projects/speechphone/

Lab: MERL Cambridge Systems
Project Type: Advanced Development

Voicelib: Voice recognition for embedded systems



Voicelib is a modular C-language implementation of various voice processing and voice recognition algorithms. Voicelib is intended for rapid prototyping of various implementations of voice recognition (both discrete-utterance and limited vocabulary continuous speech) for embedded applications.

Background and Objectives: Voice recognition in embedded systems (e.g. home entertainment, in-vehicle information systems, personal communication, consumer electronics) is expected to be a major product differentiator in the upcoming decade. Voicelib is designed to allow rapid prototyping of various user interfaces and interaction styles. Voicelib is based on publicly available GPL code to minimize coding effort where possible.

Technical Discussion: Voicelib currently implements the following functions on Linux with both a C-language interface and a shell interface: voice capture via commodity soundcard, voice detection and voice gating; format conversions; FFT and inverse FFT transforms; constant-ratio, mel, and mel-cepstral rebinning; feature-vector extraction; sample reordering and tagging; sample plotting, and orthonormal matching. Modules to perform DTW, DTFW, and Viterbi matching are currently under development.

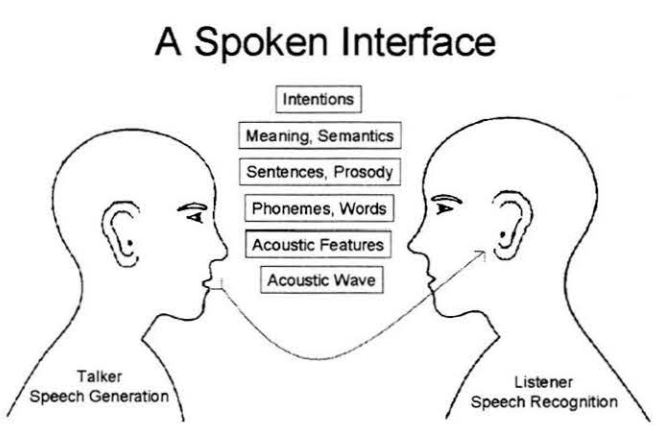
Collaboration: Currently, Voicelib is being constructed in MERL-CRL, and is being considered for use in MERL-CSL for personal communications devices.

Future Direction: We intend to expand Voicelib to have better DTW, DTFW, and Viterbi matching modules, as well as provide a library of pre-sampled speech models. If there is a need for a non-GPLed version, we are able to produce that.

Contact: William Yerazunis
http://www.merl.com/projects/voicelib/

Lab: MERL Cambridge Research
Project Type: Research

Evaluation of speech recognition engines



This project aims at implementing a standardized framework for the evaluation of speech recognition engines. This permits a completely objective evaluation of speech recognition systems for their suitability for any given task.

This framework includes the design, development and implementation of standardized interfaces and tests to measure the performance of different engines in terms of their basic recognition accuracy on standard data corpora, robustness to various degradations, and their memory and computational requirements.

Background and Objectives: At the heart of any speech-enabled application lies a speech recognition engine. The performance of the application is critically dependent on the recognition performance of this engine. The performance of speech engines differs depending on various factors the task, the operating condition and the basic engineering of the engine itself. It therefore becomes essential to evaluate all available speech recognition engines in a systematic and completely objective manner to determine which one is best suited to a given task.

The objective of this project is to design and implement a framework that can be used to characterize the performance of any speech recognition system in terms of a pre-determined set of parameters. Such characterization permits us to compare the performance of recognition systems under the constraints of the desired task in a completely objective and repeatable manner. It also enables us to identify the optimal operating point for any system.

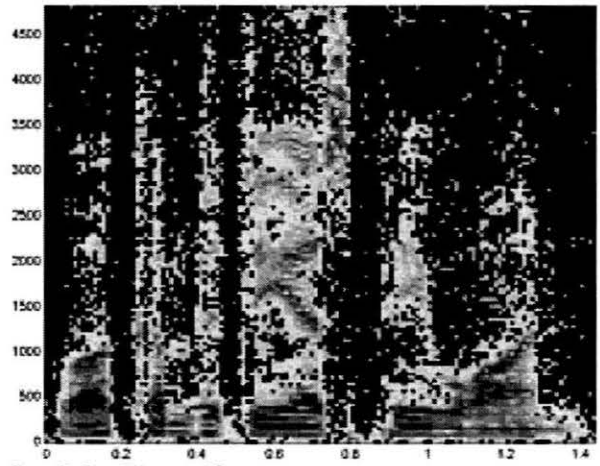
Technical Discussion: The performance of a speech recognition system is affected by many parameters, including the acoustic quality of the speech signals, any corrupting environmental influences, the complexity of the underlying task, and the memory and computational resources available. The evaluation framework therefore has to consider all of these variables. Thus, the performance of the system must be evaluated on several standardized databases such as the TI digits, Macrophone, and Broadcast News databases, as well as more realistic databases such as those collected over cellphones and in Cars. The effect of varying levels of background noises of different types must also be evaluated. In each case the performance of the system under different memory and CPU profiles must be evaluated.

Future Direction: The standardized framework will help us not merely evaluate, but also improve on any recognition systems considered, including those being developed in house and those available as open source. Consequently, this will considerably speed up the design of operating parameters and development of the speech recognition engine component of any future projects involving speech recognition.

Contact: Bent Schmidt-Nielsen
http://www.merl.com/projects/Speechengineeval/

Lab: MERL Cambridge Systems
Project Type: Advanced Development

Improving Speech Recognition with Partial Data Methods



See Color Figure 12

This project deals with the improvement of the performance of speech recognition systems in noisy environments. This is particularly essential in situations such as systems that work with far-field microphones. Noise corrupted locations of spectrographic displays of the noisy speech are identified and erased from the display. The erased regions are then reconstructed based on a priori knowledge of the statistical relationship between elements of the spectrographic display. Acoustic parameters derived from the reconstructed display are then used for recognition.

Background and Objectives: The performance of speech recognition systems degrades greatly when speech is corrupted by environmental noise. This can be a serious problem in situations where high levels of environmental noise are expected, such as in systems using far-field microphones or in automobiles. The objective of this project is to improve the performance of speech recognition systems in such situations. The algorithms developed are also applicable in other situations such as in distributed speech recognition over noisy channels where some packets may be lost in transmission.

Technical Discussion: It is expected that the effect of noise on speech signals can be better compensated for by increasing the dimensionality of the signal such that the noise-corrupted components can be better localized and isolated. The spectrogram readily presents us with such a representation. The problem of noise compensation now becomes merely one of isolating noise-corrupted regions of the spectrogram and dealing with them appropriately. There are several choices of approach once the noisy regions are identified, e.g. ignoring them during recognition or reconstructing them from the clean regions. We choose the latter option, as this does not require modification of the recognizer, enabling us to use any commercial recognition system that permits usage of externally computed front-end features. Identifying noisy regions of the spectrogram, however, remains a difficult problem. Several methods have been attempted with varying success.

Future Direction: Better identification of noisy regions is essential to make the technology completely viable. Further, it is essential that the compensation algorithm be implementable in an efficient manner to enable its usage in practical speech recognition systems.

Contact: Bhiksha Raj
http://www.merl.com/projects/missingfeature/

Lab: MERL Cambridge Research
Project Type: Research

Color Figures

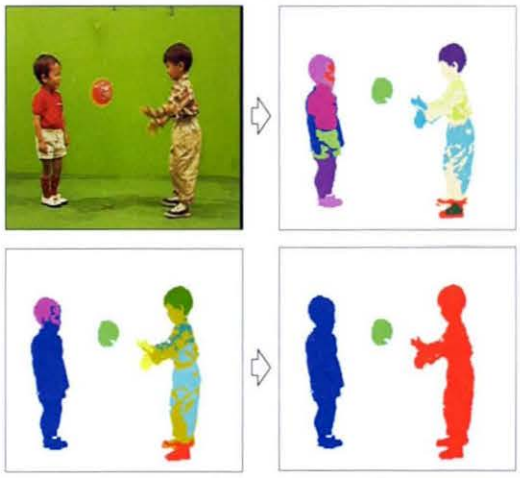


Figure 1 – see page 53

Figure 2 – see page 54

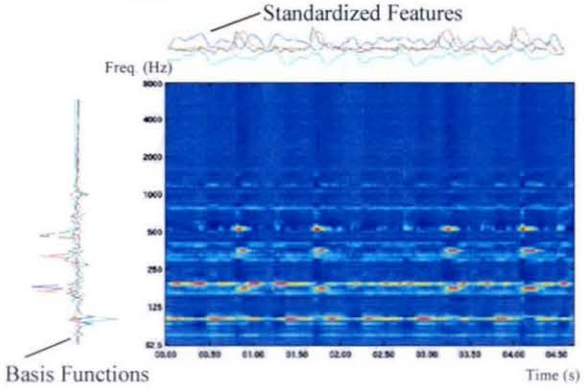


Figure 3 – see page 58

Figure 4 – see page 64

Figure 5 – see page 72

Figure 6 – see page 74

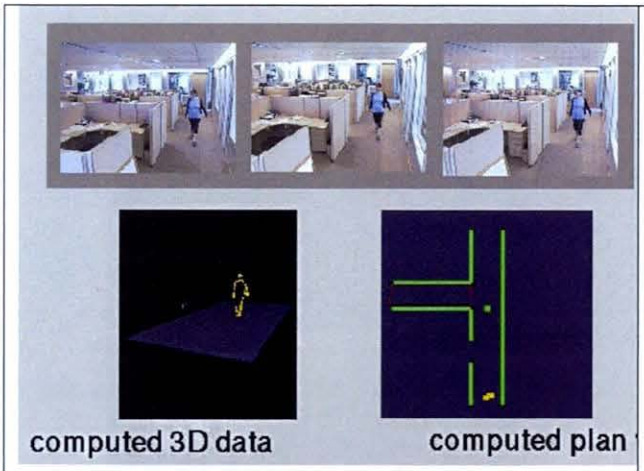


Figure 7 – see page 79

Figure 8 – see page 91

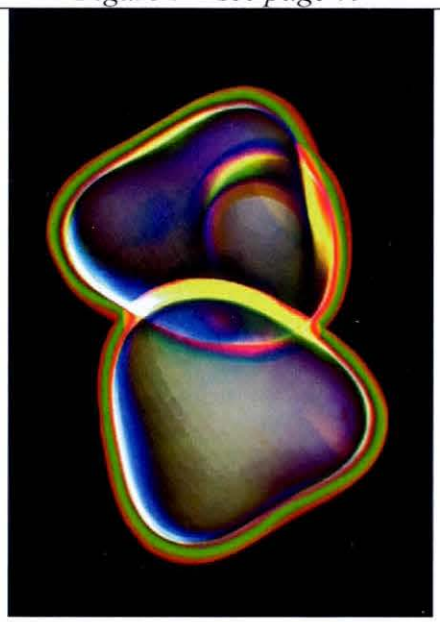


Figure 9 – see page 94

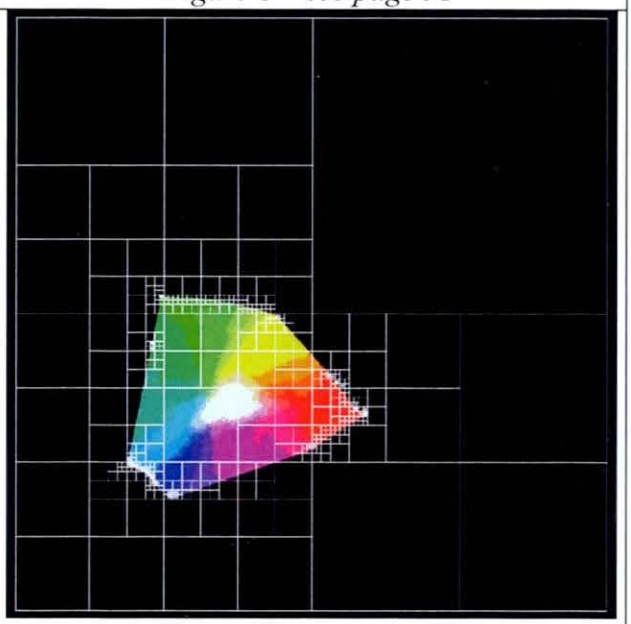


Figure 10 – see page 98

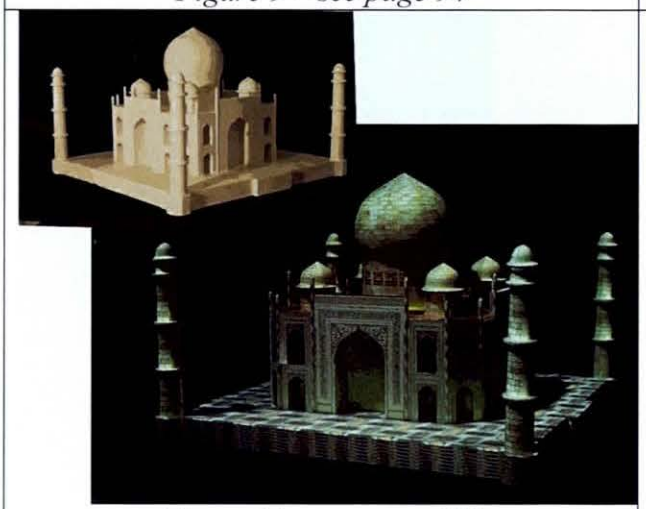


Figure 11 – see page 101

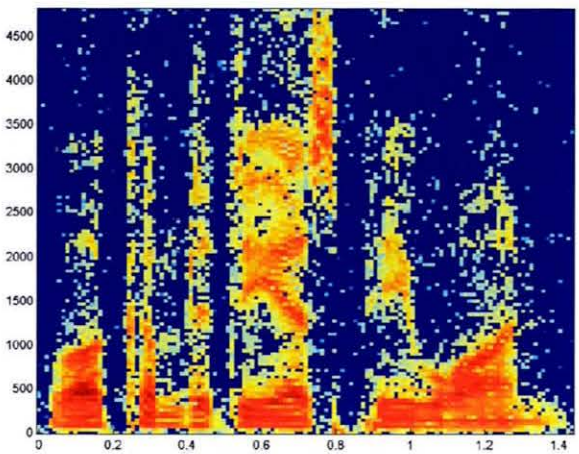


Figure 12 – see page 127