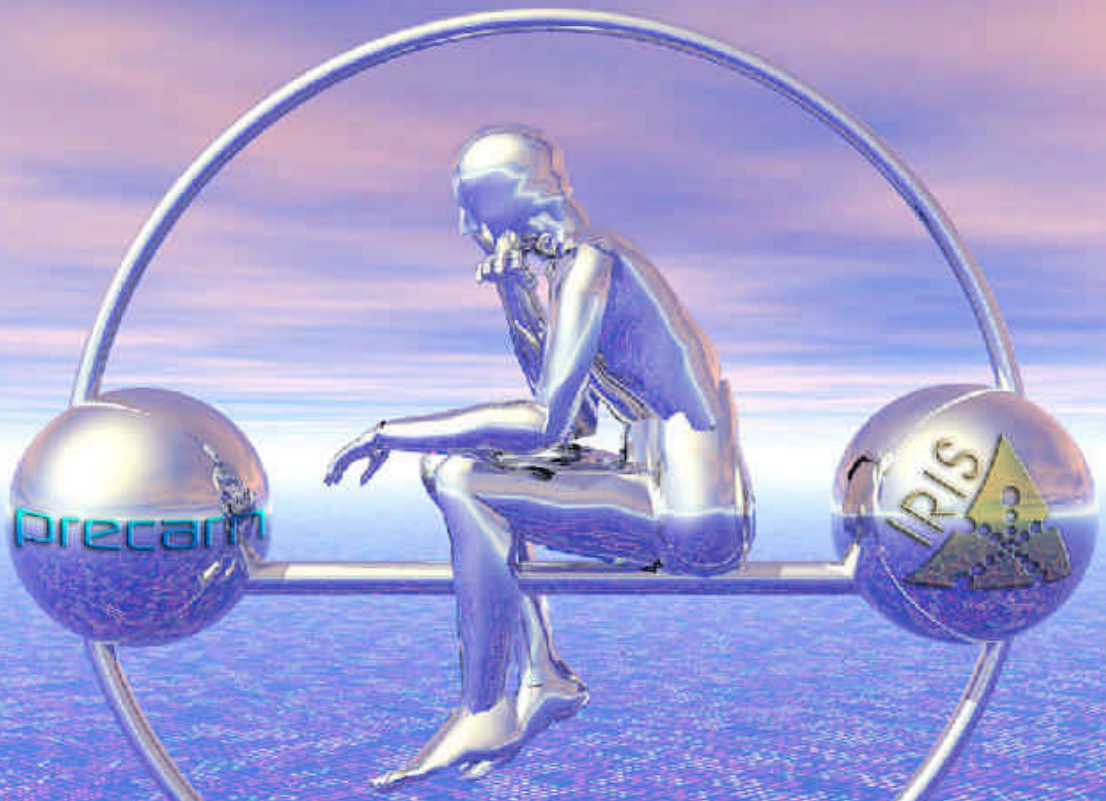
13th Annual Canadian Conference on Intelligent Systems

IS 2003

13e Conférence canadienne annuelle sur les systèmes intelligents

June 8 –10, 2003 · Halifax · Du 8 au 10 juin 2003



POSTER ABSTRACTS / COMMUNICATIONS AFFICHÉES

IRIS is a member of the Networks of Centres of Excellence (NCE) Program
IRIS est un membre du programme des Réseaux de centres d'excellence (RCE)

Table of Contents

Precarn Scholars Posters

Researcher Posters

The Generalized MDL Approach for Summarization and Its Applications

Xiaodong Zhou
Department of Computer Science
University of British Columbia

There are many applications in data analysis where we want to identify some data of interest. For example, in OLAP, an analysis query needs to find interesting records with aggregate sales greater than a threshold. The aggregation is based on the product, location and season. Usually such applications just return a set of individual records. Those individual records cannot provide enough information for the user. In the example above, although the analysis query only returns a small portion of the whole data set, the output can still contain thousands of records. It is still too huge to understand. Thus, a concise description is more preferable than a set of individual data.

Minimum Description Length (MDL) is a well-known principle to find such descriptions. It has been applied in many research areas, e.g., data compression, image segmentation, decision tree construction and so on. In our project, we extend the MDL principle to the Generalized MDL (GMDL) principle by introducing "do not care" data[1]. In GMDL, the whole data set is categorized into 3 classes: "interesting", "uninteresting", and "do not care". We include some part of the "do not care" data into our description of "interesting" ones, as long as they can contribute to a shorter description. Generally speaking, the introduction of "do not care" data can lead to a shorter description than that of MDL.

We study the summarization problem in both the spatial case and the hierarchical case. We apply the MDL and GMDL principles to both cases and develop several algorithms. The experimental results show the effectiveness of the GMDL principle and our algorithms in both run time and compression performance.

[1] Laks V.S. Lakshmanan, Raymond T. Ng, Christine Xing Wang, Xiaodong Zhou, Theodore J. Johnson "The Generalized MDL Approach for Summarization" VLDB Aug. 2002, Hong Kong, China.

Description and Analysis of Selected Hockey Game Situations

Fahong Li
Department of Computer Science
University of British Columbia
Name of Supervisor: Robert J. Woodham

Hockey coaches and players need help to improve play in game situations. We develop a system to represent and reason about hockey using as input actual player motion trajectory data tracked from game video and supported by knowledge of hockey strategy, game context and specific player profiles. We take into account five important roles that any representation plays, namely, as a surrogate, as a set of ontological commitments, as a theory of intelligent reasoning, as a medium of efficient computation, and as a medium of human expression. We take as input raw player motion trajectory data consisting of space-time point sequences of forward/backward skating registered to rink coordinates. This is augmented by knowledge of puck possession and of specific player attributes (e.g., shoots left, shoots right). We focus on the analysis of several clearly identifiable situations, such as a 2-on-1 offensive attack, defensive zone breakouts and power play shots from the point. We describe what happened in each identified situation, assess the outcome, estimate when and where key play choices were made and attempt to predict whether better options were available in the given situation. The framework is extensible both to allow the inclusion of additional hockey situations/strategies and to adapt to more general spatiotemporal data analysis.

Learning Eigen-Functions: Links with Spectral Clustering and Kernel PCA

Jean-François Paiement, Master Degree Student
Laboratoire d'Informatique des Systèmes Adaptatifs
Université de Montréal, Montréal, H3C 3J7

Spectral clustering can give very impressive results and has attracted much interest in the last few years. It is based on two main steps: first embedding the data points in a space in which clusters are more "obvious" (using the eigenvectors of a Gram matrix) , a space in which the structure of the data is revealed, and then applying a classical clustering algorithm such as K-means, as in [1]. What appears almost magical is the way in which sets of points that are on different highly non-linear manifolds can get mapped (in the above first step) to almost linear subspaces (different for each of these manifolds.)

One problem is that the procedure is highly sensitive to the choice of the kernel, for example to the choice of the spread with a Gaussian kernel. Another is that the procedure provides an embedding for the training points, not for new points. A very similar method for dimensionality reduction has been proposed in [2], based on so-called Laplacian eigenmaps. Kernel PCA is another unsupervised learning method that was proposed earlier and that is based on the simple idea of performing Principal Components Analysis in the feature space of a kernel [3].

We show a direct equivalence between spectral clustering and kernel PCA, and how both are special cases of a more general learning problem, that of learning the principal eigenfunctions of a kernel, when the functions are from a Hilbert space whose inner product is defined with respect to a density model.

This defines a natural mapping for new data points, for methods that only provided an embedding. The analysis also suggests new approaches to unsupervised learning in which abstractions such as manifolds and clusters that represent the main features of the data density are extracted. We suggest that abstractions discovered at one level can be used to build higher-level abstractions.

[1] Ng, A.Y., Jordan, M.I., and Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. In NIPS 14, Cambridge, MA. MIT Press.

[2] Belkin, M. and Niyogi, P. (2002). Laplacian eigenmaps and spectral techniques for embedding and clustering. In NIPS 14, Cambridge, MA. MIT Press.

[3] Schölkopf, B., Smola, A.J., and Müller, K.-R. (1996) Nonlinear component analysis as a kernel eigenvalue problem. Technical Report 44, Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

Exploiting a Linguistics Decomposition for Vision Based Gesture Recognition

Konstantinos G. Derpanis, Richard P. Wildes and John K. Tsotsos
Department of Computer Science, Centre for Vision Research
York University, Toronto, Ontario
E-mail: {kosta,wildes,tsotsos}@cs.yorku.ca

With current computing power, more natural forms of human/machine interaction than previously feasible are within reach. Due to their expressive power, hand gestures are an interesting form of interaction and are the focus of this presentation. Computer vision provides methods to acquire and recognize gestures while being minimally intrusive, this is a key requirement for a gesture recognition system. Additional requirements of a gesture recognition system are: recognition accuracy, real-time performance, and scalability to encompass sizable vocabularies. To ensure scalability, we appeal to linguists, specifically American Sign Language (ASL) linguistics, to define a finite set of contrastive primitives, termed phonemes that can be combined to represent an arbitrarily sized vocabulary. Currently, we focus on the extraction and classification of movement phonemes.

Our approach to gesture recognition centres around two main ideas. First, linguistic theory can be used to define a representational substrate that systematically decomposes complex gestures into primitive components. Second, it is desirable to recover the primitives from data that is acquired in as minimally constrained a fashion as possible, e.g., with a standard video camera.

We take the input to our system to be a temporal sequence of images that depicts a single movement phoneme. We also assume that the region corresponding to the hand in the first frame has been delineated by a human operator. The output of our system is a classification of the depicted gesture as arising from one of the primitive movements, irrespective of other considerations (i.e., irrespective of hand position and shape). To affect the recognition, a robust, hierarchical, affine motion estimator is applied to regions of interest defined by skin colour on a frame-to-frame basis. The resulting time series of affine parameters are individually accumulated across the sequence to yield a signature that is compared to prototypical signatures using a nearest neighbour classifier. The prototypical signatures result from a derivation of kinematic features from the apparent motion of the hand while undergoing each of the primitive movements.

The system achieves a recognition accuracy of 97.13%, with the software {unoptimized C++) based tracker running at 8 frames/sec; the time consumed by the recognition step is negligible. The hardware configuration includes: a PC workstation with a Pentium 4 processor running at 2.1 GHz and a Unibrain Fire-i 400 industrial FireWire camera.

FPGA-Driven Real-Time Person Detection: Firmware as a Processing Medium

Pierre-Olivier Laprise, Masters Degree Student
Center for Intelligent Machines
McGill University, Montreal, H3A 2A7

Real-time person detection is an important first step in many higher level vision algorithms, such as tracking. It can also be used to detect regions of interest in an image, which can then be targeted for more in depth processing that would be unfeasible in full-sized images. However, most detection algorithms are computationally expensive, and unfeasible for real-time implementation. Additionally, an algorithm was required which could be implemented in a Parallel Distributed Camera Array (PDCA) environment, and therefore implementable on a Field Programmable Gate Array (FPGA).

The algorithm that was implemented in this project, described in [1], uses cascaded classifiers to search through windows of varying size and position in the image. The classifiers look for the presence of certain weighted features in each window, discards the windows which have too few matches, and pass on the rest to the next stage. A window is considered a match if it passes through all stages. To save on computation, the features, which are sums of rectangular regions in the image, are calculated from an integral image instead of the standard intensity image.

The image source being JPEG encoded, it was necessary to first decompress the image. However, the image is only partially decoded, with an approximate integral image being computed directly from the Discrete Cosine Transform (DCT) coefficients. The entire algorithm was coded using a Hardware Description Language (HDL) for implementation on Xilinx's Virtex-II FPGA. File I/O and global system control will be handled by software running from a microBlaze firmaware processor, mapped into the same FPGA as the algorithm firmware.

The entire system will be tested using a V2MB1000 development board from Insight-Memec. All processing will be performed entirely on the FPGA, with external communications being limited to file transfers to the board, and an indication of the location of positive matches coming from the board.

[1] Paul Viola and Michael Jones. Robust Real-Time Object Detection. Second International Workshop on Statistical and Computational Theories of Vision - Modelling, Learnig, Computing, and Sampling -- July 2001

Fusion of Trinocular Stereo and Inertial Data for Underwater Robotic Navigation

Andrew Hogue and Dr. Michael Jenkin
email: hogue,jenkin@cs.yorku.ca

Mobile aquatic robots have sparked interests in many fields including biology, archeology, environmental science, and marine science. Many applications for such vehicles exist including exploration, environmental/biological monitoring, and event mapping. The aquatic environment presents many challenging and difficult problems for robotics, not the least of which are issues related to sensing. Due to the inherent physical properties of the marine environment, vision systems for aquatic robots must cope with a host of geometrical distortions, colour distortions, dynamic lighting conditions and suspended particles (known as 'marine snow'). The unique nature of the aquatic environment invalidates many of the assumptions of classic vision algorithms, and solutions to even simple problems -- such as stereo surface recovery in the presence of suspended marine particles -- are not yet known. This work provides preliminary results obtained with a novel Trinocular vision/inertial sensor system.

Inertial navigation systems (INS) have found applications in various autonomous systems for the determination of the relative pose of a vehicle over time. INS make measurements of the physical forces applied to them and thus under normal conditions they provide independent measurements of relative motion. Unfortunately these systems drift, and thus typically they are employed with some secondary sensing system in order to counteract this effect.

Realtime trinocular stereo sensors permit the recovery of 3D surfaces, however the main problem is with registration between successive views of the scene. Integrating an inertial 6DOF navigation system with a trinocular stereo sensor simplifies the registration process by providing relative motion information between frames. With this initial estimate of the camera pose, fewer features must be used to refine the registration to the global coordinate system.

Building Adaptive Agents: An Application In Computer Games

Xiaomeng Wu, Masters Degree Student
Dept. of Computing Science
University of Alberta, Edmonton, T6G 2E8

More recently, interactive computer game is an emerging application area to provide a rich environment for research on human-level AI: real-time response, adaptive behavior in changing world, cooperation and communication, reasoning, planning and learning, etc. In game, human-level AI is used to control single agent, provide strategic decision to multiple agents, dynamically change parameters to make the game more challenging, or produce in-time and helpful hint for human players, etc. Compared to military simulation, robotics, it offers an inexpensive, reliable and accessible research environment.

The present project aims at building adaptive intelligent agents, which are capable of responding to other agents and/or its environment to some degree. More precisely, the task an adaptive agent must perform can be divided into four different aspects: reaction, reasoning, learning and evolution. They can be used singly or in any combination. Specifically, building learning ability for adaptive agents is a broad field with application in robotics, education, e-commerce, military, WWW, network, etc. It is also much related to many other AI research areas, such as plan recognition, opponent modeling, knowledge acquisition.

Machine learning has been used in classical board games and gained success [1]. In comparison, interactive computer games provide a more realistic physical world for learning in which perception, information-processing and action-taking is much more close to human world. For this purpose, a layered architecture is designed to model the knowledge representation and internal behavior structure for an agent. In this architecture, the agents are able to represent the other agents and the environment to facilitate recognition and reasoning by a layered model. Knowledge and behavior can be assembled, decomposed and reasoned by meta units, which will help agents to map the real world more efficiently, develop learning through building of feedback mechanism, and make the regeneration of their own dynamical structure possible and flexible according to the changing environment.

The recognition system, which takes massive raw data perceived from sensor as input and output the state of the world in a relatively abstract level using high level description, is a very important part for all the layers above. Current project is using Hidden Markov Model (HMM) to autonomously identify the agent's behavior. HMM has been successfully used in recognition tasks in many areas [2]: speech, handwriting, vision, etc. The hidden states in HMM can be mapped naturally into the unobservable mental states of agents, and the set of observations corresponds to the physical state of the agents. It provides a modeled abstraction of the potentially complex behavior and has been tested in robotics domain [3]. In the more realistic computer soccer game, much more complex inputs and behaviors are involved. Therefore, Hierarchical Hidden Markov Model (HHMM) is implemented for the multiagent behavior recognition, and parameters of the models are also tuned by training process.

[1] Johannes Furnkranz, “Machine Learning in Games: A Survey", Machines that Learn to Play Games, Nova Scientific Publishers, Chapter 2,pp.11-59, Huntington, NY, 2001

[2] Olivier Cappe, “Ten years of HMMs", 2001

[3] Kwun Han, Manuela Veloso, “Automated Robot Behavior Recognition Applied to Robotic Soccer", Robotics Research: the Ninth International Symposium, IJCAI 1999

Adaptive Characters in Commercial Games

Jonathan Newton
10028-106 Ave #406
Edmonton, AB
T5H 4A3
(780) 425-2551

This research investigates one application of machine learning to commercial games. Using role-playing games as a test bed, a system that automatically adjusts difficulty levels for a player could be created. This requires learning from the previous play of the human player. Once the play style has been learned for a specific player the game could then be adapted so that the difficulty level perceived by this player and the difficulty level intended by the developers are more correlated.

Pathfinding in Computer Games

Peter Yap, PhD student
University of Alberta
Department of Computer Science
Edmonton, Canada

Pathfinding has applications in many domains, including the multi-billion dollar games industry. Pathfinding in this domain is atypical in that we have interesting real-time constraints. A search must be completed in the milliseconds with limited memory, quite often the complete path could not be found in this time frame. Pathfinding algorithms must be quite clever to deal with these constraints. The key criterias for a good pathfinding algorithm in a computer game are that 1) the path looks realistic 2) the path looks somewhat optimal and 3) the path can be found quickly on a below average computer. While the choice of implementation of a pathfinding algorithm is important, the underlying topology on which the search is conducted on is of no lesser importance.

In this poster, we will show how pathfinding is done in computer games in general, and why the problem is so difficult. Using examples of real computer games, we will present our progress so far in this field. Much of our research is done with the collaboration of computer game companies, including BioWare and Electronic Arts.

Modeling Needle Insertion in the Thoracic Cavity Using Finite Elements and a Single Transverse CT Slice

Danny French, Masters of Applied Science Student
Robotics and Control Laboratory,
University of British Columbia, BC

Inserting long needles into the thoracic cavity is frequently required for diagnosis (i.e. obtaining biopsy samples) and treatment (i.e. brachytherapy) of disease in the lungs. To effectively direct a needle to a small nodule physicians require real-time visual feedback, most often provided to them using fluoroscopy. Experienced physicians significantly improve their technique by using kinesthetic feedback to guide the needle to the region of interest. Kinesthetic feedback is experienced through forces and torques on the needle shaft. However, inexperienced physicians have great difficulty interpreting and using force feedback, and consequently often miss the target. This results in bleeding in the region of interest, thus making it difficult to visually distinguish small details in the real-time image. Moreover, because fluoroscopy is used, physicians are exposed to harmful radiation throughout the procedure.

To help less experienced physicians learn to use force feedback for controlling needle trajectory during long needle insertion into thoracic cavity, the present work proposes to model the forces and torques on the needle using a computed tomography (CT) slice of the region of interest together with a finite element model of the tissues. As part of this work, an algorithm to segment a transverse CT slice of the thoracic cavity will be implemented. The algorithm will use thresholding to identify the lungs and active contours to identify boundaries between soft tissues (i.e. fat and muscle).

After the different anatomical regions are segmented, elasticity parameters will be assigned to each region. Next, the 2-dimensional (2D) image will be meshed into homogeneous elements. These meshes will be designed such that nodes lie on the boundaries of inhomogeneities in the image. Then, a model similar to the one presented in [1] will be used to model the forces and torques on the needle as it travels to a desired target (such as a lung nodule).

Modeling forces and torques on a needle passing through the thoracic cavity can be used to preoperatively plan an optimal needle trajectory, to provide force feedback to the physician in a teleoperative environment, and to accurately simulate procedures for training purposes.

[1] "Needle Insertion Modeling for the Interactive Simulation of Percutaneous Procedures," S. P. DiMaio and S. E. Salcudean. Proceedings of Medical Image Computing and Computer-Assisted Intervention, Tokyo, September 2002.

Hand Detection and Tracking in an Active Vision System

Yuliang Zhu and John K. Tsotsos
Department of Computer Science
Centre for Vision Research
York University, Toronto, Ontario
Email: {ylzhu,tsotsos}@cs.yorku.ca

As the impact of modern computer systems on our every day life increases, human-computer interaction (HCI) has become more and more important in our daily lives. In fact, as the computing communication, and display technologies progress, the existing HCI techniques, such as mice and keyboards, limit the speed and naturalness of our interaction and may become a bottleneck in the effective usage of computers. HCI research hopes to permit interaction in a natural manner by using gesture and gaze. One domain of application is video-conferencing. In most current teleconferencing or distant learning systems, the camera is fixed or is controlled by an operator. However, in a natural communication between people, gestures, facial expression and body language also play an important role. We use a number of methods to direct the visual attention of those with whom we interact. One very common tool is 'to point' with a figure to items of interest.

Motivated by the above ideas, a hand tracker was implemented, which detects and tracks a hand in a pointing gesture by using the Condensation algorithm. The background may be highly cluttered, and the stereo cameras move actively. The tracker propagates the density of the hand states (translation, rotation and scaling) by sampling the state space and applying the motion model, followed by measuring the hypotheses of the hand contour on the skin color map of the raw images. By utilizing the parameters of the camera system, the 3D orientation of the hand is calculated using the epipolar geometry.

The system consists a robotically controlled binocular head, connected to a Dual Pentium II PC platform. Each color camera mounted on the robot head can be controlled independently. Two sets of video are captured by Imaging Technologies S-Video frame grabber, with resolution of 512x480 pixels and color depth of 24 bits. The cameras can be used independently or as a stereo pair. There are 4 mechanical degrees of freedom: head tilt and pan for head and each camera. The server computer directly connects with the motors and cameras (control part) of the robotic head through motor control cards and serial port, respectively. The client computer, where the application is running, has two video input from the cameras and can send TCP/IP packet to the server to set and get the parameters of the motors and cameras, so that the head is controlled by the application.

Constraint-based Design of Ainia, a Robotic Kicker

Pinar Muyan, M.Sc. Student
Department of Computer Science,
University of British Columbia, Vancouver, BC, V6T 1Z4

The Constraint Based Agent (CBA) framework (1) is one of the approaches proposed for the study of hybrid systems. The CBA approach differs from other methods by advocating the following arguments: 1) Hybrid systems should be modeled using formal methods. These methods should state how the system is composed and how it works. They should specify requirements that indicate what the system should do overall and verify that the behaviors of the system satisfies these requirements specifications. 2) The approach should have a unitary framework that supports both discrete and continuous time/domain structures. 3) Since hybrid systems are complex systems with multiple components, the model should support hierarchy and modularity. 4) The new model for hybrid systems should be at least as powerful as existing models. 5) Agent controllers should be specified and designed as online constraint satisfiers.

Many studies have been conducted within the CBA framework which propose the CBA approach as a design framework for situated agent design. However, these studies utilized a limited version of the CBA model, which did not have prioritized constraints and arbiter structures. Additionally, only two of these systems were implemented with Constraint Nets in Java (CNJ), a visual programming environment for constraint nets (2). Moreover, the systems that were designed and simulated with CNJ were not implemented or tested in the real world.

This study proposes that the extended CBA framework with prioritized constraints is an effective methodology for specifying, designing, simulating and building situated agents. For this purpose, first a robot called Ainia (an Amazon warrior whose name means ``swiftness''), that repeatedly finds, tracks, chases and kicks a ball in the field has been designed and simulated in CNJ. After modeling the controller, the body and the environment as separate modules and creating an animation of the system under CNJ, the controller module of the Ainia is then used to control a physical robot. The body and environment modules are replaced by the physical robot plant and the external real world, respectively. Since the behavior of the real robot satisfies the constraint-based requirement specifications while demanding few changes in the controller, this study provides evidence that CBA approach is an effective framework for situated agent design and construction. It also supports the claim that CNJ is an effective tool for designing robots in the real world.

A. K. Mackworth: 2000, 'Constraint-Based Agents: The ABCs of CBAs'. In Proc. 6th Int. Conf. on Principles and Practice of Constraint Programming - CP 2000, Singapore, LNCS Vol. 1894, pages 1-10, Springer Verlag.

F. Song, and A. K. Mackworth: 2002, 'CNJ: A Visual Programming Environment for Constraint Nets'. In Proc. of AI, Simulation and Planning in High Autonomy Systems -AIS 2002, Lisbon, Portugal, pages 131-135.

Haptic Guides: Providing Usable Force Feedback Information in Time Critical Tasks

Ben Forsyth, MSc Student
Department of Computer Science, UBC
Vancouver BC V6T 1Z4

Cars of the future will know much more about their surrounding environment. We are looking for ways to use this information to make driving safer and easier. We are investigating control algorithms that use this information to provide driving suggestions via force feedback through the steering wheel. We have developed some prototype intelligent control methods that attempt to take into account how users reflexively react to guiding forces in time critical situations. To reduce the size of the problem we are investigating we will not be performing experiments on actual driving tasks. Instead, we will study how our control algorithms perform at guiding users along a 2D path and through a 2D channel (such as the lane you drive in).

There are two main types of problems in developing effective haptic lateral guidance: control methods to generate forces and the user's reaction to these forces. When developing active force feedback applications you must be careful when designing control methods for the force feedback device. Simple control methods can lead to instabilities such as oscillations from strong corrective forces when going around a corner or approaching a side of a channel too quickly. The reaction forces in these situations can induce oscillations by pushing the user past the desired position and into another configuration that creates a strong reaction force and so on.

Having human users in the control loop creates other design challenges. In time critical tasks such as driving the user does not have time to think about what the forces they are feeling mean. Not only does the user need to understand what the forces mean without thinking about them, but the forces must also be transparent enough that the user does not reflexively fight the forces presented to them. Assuming that the understanding and transparency problems can be resolved there is still the possibility that the user may not want to do what the system thinks that they should do. Therefore, the system should be compliant enough to allow the user to perform tasks other than what the system is suggesting.

We are currently investigating two different approaches to address the problems stated above. One approach is to extend work on generating realistic path following behaviors for autonomous creatures to include a human in the loop [2]. These methods predict the user's location in the future and will hopefully avoid control oscillations and frightening the user by providing gentle guiding forces before the user reaches a trouble location. Our other approach at solving the problems above involve extending existing work on using haptics for path guidance in surgical applications [1]. This work is good at providing path guidance as well as being compliant to off-path motions, but it is not predictive.

We will use SensAble Technology's PHANToM force feedback device to implement our control methods. User studies will be performed to compare our proposed control algorithms to simple control methods such as a spring and damper for path guidance and a potential field for channel guidance.

[1] A. Bettini, S. Lang, A. Okamura, and G. Hager, “Vision Assisted Control for Manipulation Using Virtual Fixtures: Experiments at Macro and Micro Scales,” Proceedings of the IEEE International Conference on Robotics and Automation, 2002, pp. 3354-3361.

[2] C. Reynolds. Steering Behaviors for Autonomous Characters, Game Developers Conference 99. http://www.red3d.com/cwr/steer/

Stereo Camera Measurement of Surfaces for Ultrasound Applications

Giselle Flaccavento, Masters of Applied Science Degree Student
Robotics and Control Laboratory, Electrical and Computer Engineering
University of British Columbia, BC, V6T 1Z4
Supervisors: Dr. P. Lawrence, Dr. R. Rohling

In recent years, ultrasound has become an increasingly popular medical imaging acquisition technique due to its low cost and non-invasive nature. New techniques, improvements, and applications of ultrasound are among current popular research topics. The goal of this project is to develop a computer vision system used for tracking during an ultrasound exam. The suggested system can be applied to both tracking surface contours of the patient and tracking the ultrasound transducer probe. By tracking the probe with respect to the patient's skin, the need to secure the patient to a fixed coordinate system is removed. The technique also eliminates the need for the patient to hold their breath during an ultrasound exam. Movement during the ultrasound exam can be due to respiration, deformation of the surface caused by probe pressure, and small movements of the patient. The system tracks the surface of the patient's skin throughout the ultrasound scan. Using images produced with digital stereo cameras, the 3D location of the surface can be tracked and recorded. The surface is partitioned into sections. A triangulation algorithm is used to determine the 3D location of each partition of the surface. A best fit for each portion of the surface is then determined. Using the set of points for each portion of the surface has the advantage of improved accuracy compared to finding the location of each point separately. One method of creating a 3D ultrasound model requires that 2D ultrasound images are stacked. The probe positional information with respect to the patient's skin when these 2D images are recorded provides the necessary information about the relationship between these images. A panoramic ultrasound image requires registration of multiple ultrasound images. The positional information about the probe and patient's skin is required in order to perform the stitching correctly. This system can also be useful with respect to surgical planning. The position of pre-operative ultrasound scans can be used in order to align the ultrasound images with the patient during surgery. Registration of ultrasound images with other medical image modalities is another use for the positional information provided by this system.

Haptic Feedback Using Local Models Of Interaction

Daniela Constantinescu, Ph.D. Candidate
Robotics and Controls Laboratory
University of British Columbia, Vancouver, V6T 1Z4

To date, a major obstacle to the wide-spread use of haptic devices is the lack of a technique to allow stable interfacing of a device to an arbitrary virtual environment simulation while providing realistic force feedback. Such a technique is challenging to develop because stability requires fixed force "refresh" rates of the order of hundreds of Hz to be provided to the user's hand regardless of the rendering frequency of the virtual environment (typically of the order of tens of Hz for physically-based rigid body simulations and possibly lower). The present project proposes to meet the need for stability and the need for a high haptic rate by decoupling the force control loop from the simulation through a local model of interaction. The model is suitable for adding realistic force feedback to the manipulation of both rigid objects and linkages.

In the proposed local model, the fixed haptic rates are achieved by limiting the interaction between the object manipulated by the user (the virtual tool) and the virtual environment to the interaction between the virtual tool and neighboring objects. The model comprises a dynamic proxy of the virtual tool together with constraints imposed on the virtual tool motion by neighboring objects. The quality of the approximation is maintained locally by updating the model at each step of the simulation. Discontinuities in the local geometry at updates may lead to undesirable discontinuities in the forces applied on the user's hand. Such discontinuities are avoided using local proxy deformation and prediction of constraints [1].

Two techniques are used in the proposed local model to improve contact stability and the perception of rigidity: (i) a local force model extending that in [2] that combines collision impulses upon contact with penalty forces during contact; and (ii) local proxy dynamics. Collision resolution is performed using an extension of Newton's restitution hypothesis that accounts for linkage constraints and multiple simultaneous collisions. Local proxy dynamics allow various coordination techniques to be used between the proxy and the device. Therefore, they allow the realism of the haptic feedback to be improved not only through an improved local model of force computation, but also through improved device control.

Haptic manipulation of linkages is incorporated in the proposed local model by defining the virtual tool as the entire linkage if the user holds one of its links and by sending neighboring contacts of all links to the local model. Motion constraints imposed on the user by the linkage topology are encapsulated in the local proxy dynamics.

[1] D. Constantinescu, S. E. Salcudean, E. A. Croft, "Local Interaction Models for Haptic Rendering of Rigid Environments", Proc. 2nd IFAC Conf Mechatronic Syst, pp. 553-558, Berkeley, Ca, 2002.
[2] S. E. Salcudean and T. Vlaar, "On the Emulation of Stiff Walls and Static Friction with a Magnetically Levitated Input/Output Device", J Dyn Syst Meas Contr, vol. 119, pp.127-132, 1997.

Measurement-Based Deep Venous Thrombosis Screening System

Julian Guerrero, Ph.D. Student,
Robotics and Control Laboratory,
University of British Columbia, Vancouver, V6T 1Z4

It is known that deep venous thrombosis (DVT) is a disease that affects the venous system, where blood clots are formed and obstruct the venous pathways, which may cause serious consequences such as pulmonary embolism (PE). The overall mortality of untreated DVT has been reported at approximately 2.5%[1]. Screening using compression ultrasound (CUS) is a routine procedure currently performed in many health care facilities to detect DVT. A technician scans a patient to determine whether a thrombus is present in the patient's venous system by performing compression exams (compression-release cycles) at different locations and identifying incompressible vein segments. However, these examinations rely on the expertise of the examiner, may take in excess of 40 minutes and in 76% of cases require repeat scans [1], increasing overall cost. Therefore, there is a need for a system that can objectively quantify thrombi and their location through a single examination. An experimental system and interface that indicate the likelihood of deep venous thrombosis using objective measures was developed, based on the well-established compression ultrasound examination. Ultrasound image data is processed using modified Star-Kalman algorithm in order to obtain a measure of the transverse vessel area. The force applied to the patient with the probe is also measured, as well as the probe location using a sensorized ultrasound probe. In this manner a vessel segment is assessed for compression. A user interface displays the acquired information using a 3-D representation of the scanned vessel. The objective measures that indicate the likelihood of deep venous thrombosis have undergone initial validation, and preliminary tests on healthy subjects have been performed. Results were also obtained for a tissue mimic or phantom developed for testing and validating the system. To date, we have had very promising results.

REFERENCES

[1] Nicolas Perone, Henri Bounameaux, and Arnaud Perrier. Comparison Of Four Strategies For Diagnosing Deep Vein Thrombosis: A Cost-Effectiveness Analysis. The American Journal of Medicine, 110(1):33{40, January 2001.

Experimental Validation of the Analysis of Mechanism Compliance

Patricia McAllister, Ph.D. Student
Department of Mechanical Engineering
Queen's University, Ontario, K7L 3N6

Determination of the spatial compliance or stiffness of a mechanical system by conventional means relies on the accurate knowledge of the kinematic geometry, and the mechanical stiffness properties, of both the rigid and elastic components of the system. For systems such as biological joints, accurate geometric or stiffness information is rarely available. In these cases, the elastic behaviour of the mechanism cannot be characterized by conventional means without making assumptions regarding the underlying geometry of the system; these assumptions can lead to inaccuracies in the resulting model. For biological systems, the stiffness models frequently omit several degrees of freedom, thereby disregarding the potential existence of coupling between motions [1]. Other experimental methods of determining the stiffness of biological joints require that axes of motion of the joint be aligned with directions of applied force [2]. These assumptions can result in experimentally finding an artificial stiffness that is, in fact, due to misalignment of axes of motion; this in turn can lead to the loss of substantial information about the existence of coupling between directions of motion. We have developed a method to experimentally resolve the characteristic compliance matrix in all six degrees of freedom (DoFs) for a passive mechanism, based on the measured elastic response to applied loads.

This project involves the experimental validation of a screw theory-based compliance characterization from quasi-static data. Experimental validation of compliance characterization from experimental data was performed by obtaining quasi-static data for four mechanisms that had well determined geometry. The four mechanisms were: one stiff revolute joint (1 DoF), one stiff prismatic joint (1 DoF); one stiff universal joint (2 DoFs + "loosely toleranced" directions); and one stiff cylindrical joint (2 DoFs +"loosely toleranced directions"). It was expected that the characteristic matrix for each of these systems would be ill-conditioned. Quasi-static data were acquired for each mechanism using a 6-DoFforce/torque sensor and an optical tracking system with sets of 4 infrared emitting diodes arranged into rigid targets. Paired twist (displacement) and wrench (load) data were collected for small displacements about an equilibrium position for the mechanism. The6x6 compliance matrix was calculated from over determined sets of paired twist and wrench data. The compliance assessment was repeated 30 times for each mechanism. The resulting compliance matrices were neither symmetric nor positive-definite due to the presence of noise in the data. Asymmetric, positive semi-definite (SPSD) approximation of the matrix was obtained using a method described by Higham [3].

In previous work [4] we developed a method of analysis based on the eigenscrew decomposition described by Patterson and Lipkin [5, 6].This method was used to find the basis screws, the magnitudes of the pitches, and the rotational and translational compliances associated with each of the four mechanisms. Experiments showed that a 6-DoF compliance characterization was successful in providing a qualitative description

of the number of degrees of freedom exhibited by each mechanism, and the general axes of motion associated with the degrees of freedom. The "loosely toleranced" directions were also identified by the compliance characterization. However, a quantitative analysis of the mechanism compliance was unreliable because of large variations (>25\% from mean value) in the magnitudes of compliance. Due to inherent scaling of the compliance matrix by the units, the directions of translational compliance were often determined by the dominating rotational compliance. We are now examining alternative analytical strategies to improve the experimental quantification of mechanical compliance. Current efforts are focused on obtaining a better first estimate of the symmetric, positive definite compliance matrix from experimental data by applying filters. In the next phase of the project, the experimental method will be applied to the analysis of a biological system.

[1] M.M. Panjabi, R.A. Brand Jr., A.A. White III. Three-dimensional flexibility and stiffness properties of the human thoracic spine, J. Biomechanics, 9:185--192, 1976.

[2] A.B. Zavatsky. A kinematic-freedom analysis of a flexed-knee-stance testing rig. J. Biomechanics, 30(3):277--280, 1997.

[3] N.J. Higham. Computing a nearest symmetric positive semi-definite matrix. Linear Algebra and its Applications, 103:103--118,1988.

[4] P. L. McAllister. An Eigenscrew Analysis of Mechanism Compliance. M.Sc. (Eng) Thesis, Queen's University, Kingston, Ontario, 1998.

[5] T. Patterson and H. Lipkin. A classification of robot compliance. In Cams, Gears, Robot and Mechanism Design, volume 26, PP 307--314. ASME, ASME design Technical Conferences, September 1990.

[6] T. Patterson and H. Lipkin. The structure of robot compliance. In Cams, Gears, Robot and Mechanism Design, volume 26, PP 315--322. ASME, ASME design Technical Conferences, September 1990.

High-Resolution Video Synthesis from Mixed-Resolution Video

AUTHORS: Stéphane Pelletier, Stephen Spackman and Jeremy R. Cooperstock
Electrical and Computer Engineering
McGill University
Montréal, QC

In order to generate a video frame, imaging devices accumulate photons over a 2D matrix of light sensors, whose number determines the maximum achievable resolution of the camera. Smaller sensors allow for a greater number of them to be contained in the matrix, and hence, a higher resolution can be achieved. The exposure (or integration) time of a single frame must be chosen so that each such sensor receives a sufficient number of photons to allow for a statistically accurate measure of the light intensity at its location. This is partly dependent on the surface area of the sensor. A physically smaller element requires a proportionately longer exposure time to produce a usable image, which in turn determines the maximum frame rate that can be achieved by a camera; shorter exposure times allow the video device to produce frames at a higher rate.

To obtain high-resolution video at higher frame rates than that allowed by the integration time of light sensors, we propose using special video hardware capable of simultaneously generating high-resolution frames H at frame rate h and low-resolution frames L at frame rate l. Since low-resolution frames involve the accumulation of incident photons over a larger sensor surface for each pixel and, thus, require less time to integrate than high-resolution frames, the frame rate l of the low-resolution sequence is naturally higher than h. The high- and low-resolution frames represent the same scene and are used respectively to capture high-frequency details and object motion. Our method applies an image-processing algorithm to both sequences of frames H and L in order to synthesize a high-resolution video sequence S, at high frame rate l, containing the detail of the high-resolution frames H and the motion dynamic of the low-resolution frames L. More precisely, our technique translates the pixels of the few high-resolution frames generated by the video hardware with respect to the motion observed in the low-resolution frames. Simulation results indicate that the visual quality of the high-resolution frames synthesized by the algorithm is far better than that of the corresponding low-resolution frames

Eccentricity Effects With Natural Images

Ljiljana Velisavljevic, Ph.D. Degree Student
Centre for Vision Research
York University, Ontario, M3J 1P3

Purpose: In studies of natural image perception, eccentricity effects may arise from (1) a decline in visual performance as a function of retinal eccentricity (observer effect), and/or (2) a decline in the detail or salience of the image content as a function of image eccentricity (framing effect). Here we assess the role of both factors in the rapid visual coding of natural images.

Method: A local recognition task was employed. Each trial sequence consisted of a fixation, test, mask and probe stimulus. The test and mask stimuli were randomly-selected 31x31 deg natural images, displayed for 59 and 506 ms, respectively. A black grid divided each of the images into 64 3.8 deg square blocks. The probe stimulus consisted of two blocks presented on either side of fixation, one drawn randomly from the test image, the other from a random image. The task was to identify which of the blocks was drawn from the test image. In order to independently assess the influence of observer and framing effects, the test stimuli were cropped from both central and non-central locations of larger natural images.

Results: Recognition performance for coherent natural images was found to decline significantly with eccentricity. However, this effect disappeared completely when the 64 blocks of the test image were scrambled. Regression analyses revealed that both retinal and image eccentricity were significant factors, although the magnitude of the observer effect was roughly twice that of the framing effect.

Discussion: Both spatial acuity and chromatic sensitivity decline with eccentricity. However, the absence of an eccentricity effect for scrambled images argues against these being the primary factors. Rather, the observer effect appears to be due to a higher-level facilitation in foveal processing triggered by the coherent structure of the image.

Non-Invasive Tracking Of The Mind's Eye

Ziad M. Hafed, Precarn Scholar
Center for Intelligent Machines, McGill University,
3480 University Street, Montreal, Quebec H3A 2A7

The human visual system displays tremendous versatility in its ability to effectively process scenes in our environment. Of particular interest to us is this system's ability to handle processing bottlenecks. Such bottlenecks give rise to a mechanism of 'attentional selection', or the selective processing of stimuli of interest, that is heavily relied on by the brain. One form of attentional selection commonly used by humans is that of an overt eye movement. By aligning gaze direction with a particular object of interest, the act of 'looking' at this object with an overt eye movement allows the brain to use the high resolution area of the human retina (fovea) to inspect this object. So, if one wishes to study how humans perceive their environment, tracking overt eye movements is important. However, overt eye movements are not the only form of attentional orienting. Humans also have the ability to 'look out of the corner of their eyes' or to 'look with the mind's eye' - a process described as covert attentional selection.

The ability to allocate attention covertly means that tracking of the mind's eye should augment conventional eye tracking technology in any application that involves monitoring how humans visually perceive their environment. However, while eye tracking is fairly easy to accomplish, no non-invasive 'overt' measure of covert attention seems to exist. In a recent paper [1], we suggested that microsaccadic eye movements, which are tiny eye movements that occur subconsciously during gaze fixation, may provide such a measure. In the present work, we demonstrate the use of this measure in an example application of 'attention tracking'. In this example, we analyze the patterns of covert attention shifts that humans undertake when executing simple visual tasks.

We designed a task in which trials consisted of the onset of a peripheral stimulus followed by a foveal stimulus. The stimulus onset asynchrony between the peripheral and foveal stimuli was randomly chosen from among 0, 50, 100, 150, & 500 ms. Subjects were instructed to maintain fixation and to make speeded same/different judgments on the colors of the peripheral and foveal stimuli. Eye movements were monitored, and microsaccade analysis revealed distinctive patterns of attention shifts: shifts to the peripheral stimulus and back after this stimulus' onset, followed by ones to the peripheral stimulus and back after the foveal stimulus' onset, followed finally by a shift to the peripheral stimulus that was tightly synchronized with manual response execution. This final attention shift is hard to uncover using indirect measures of covert attention. When the task involved making overt eye movements to the peripheral stimulus, the patterns of covert attention shifts occurring between these eye movements were identical to those in the fixation variant of the task. We conclude that the accessibility to covert attention shifts that microsaccade analysis allows has important implications on the study of how humans employ covert attention when interacting with their visual environment.

[1] Hafed, Z. M. & Clark, J. J. (2002). Microsaccades as an overt measure of covert attention shifts. Vision Research, Vol. 42, 2533-2545.

Acoustic Positioning for Underwater Autonomous Vehicle Localization

Hui Liu, Evangelos Milios
Faculty of Computer Science
Dalhousie University, Halifax B3L 4P7

An important problem associated with the navigation of an underwater vehicle is how to accurately localize it. We are investigating techniques to solve the problem using acoustic means. We use an array of hydrophones floating at the surface, whose absolute position can be measured via GPS. We estimate the direction of arrival at the hydrophone array of sound in the audio range from a sound source mounted on the robot. Estimation of the direction of arrival at multiple hydrophone arrays allows estimating the position of the underwater robot in 3D space. Novel array designs are considered, and an optimal solution for handling errors caused by all possible noise sources in the measurement is provided. The experiments to test the accuracy and consistency of our designs are carried out both in the air (using a scaled array model with condenser microphones) and in the water. Experimental results with real data are presented.

Context-Based Object Detection

Tina Ehtiati, Ph.D. Student
Centre for Intelligent Machines
McGill University, Montreal, H3A2A7

Experiments in human scene perception and visual search have provided increasing evidence that the human visual system uses the relationships between the environment and the objects within the environment for facilitating object detection and recognition [1]. Object-centered approaches for object detection and recognition only make use of the intrinsic object features, while in many real-world scenes, contextual information can also provide relevant information for these tasks. Our goal in this project is to demonstrate the role of contextual information in the detection of individual objects in real-world scenes. In this project contextual information of the scene is defined as a holistic representation of the scene, based on statistics of low-level features of the scene, as encoded by oriented band-pass filters [2]. A probabilistic framework is used for the formulation of the contextual influences on object detection and scene classification. The contextual information provides an estimate of the likelihood of finding an object belonging to a certain object class and its most likely positions and scales in the scene. The computation of the context-based object likelihood requires a learning stage, where the statistics of the low level features are conditioned according to the presence or absence of objects and their scales and locations in the scene. EM algorithm is used as the learning method for this purpose. At a higher level of conceptual abstraction of the scene content, an estimate of the likelihood of the scene class is provided based on the object-level likelihood estimates. In this scheme single and joint probabilities of the presence or absence of certain object classes and their relative positions and scales in the scene influence the estimate of the likelihood of the scene category. A database of 400 real-world images consisting images of four abstract scene categories has been used for this project. Four abstract scene categories: streets, parks, houses, indoors, and four object categories: vehicles, people, trees, furniture, have been used in the experiments. The results of the experiments indicate the use of contextual features for inferring information about the object classes present in the scene and forming a hypothesis about the symbolic meaning of the scene.

[1] A. Torralba, P. Sinha, “Statistical context priming for object detection”, Proceedings of the International Conference on Computer Vision, ICCV01. (pp. 763-770), Vancouver, Canada, 2001.

[2] A. Torralba, A. Oliva, “Semantic organization of scenes using discriminant structural templates”, Proceedings of the Seventh International Conference on Computer Vision, ICCV99. Vol. 2, pp. 1253-1258, Korfu, Greece, 1999.

Probabilistic Constraint Nets: A Unified Framework for Probabilistic Hybrid Systems

Robert St-Aubin, PhD Student
Laboratory for Computational Intelligence
University of British Columbia

The development of autonomous agents, such as mobile robots or software agents has generated considerable research in recent years. Robotic systems, which are usually built from a mixture of continuous (analog) and discrete (digital) components, are often referred to as hybrid dynamical systems.

The modeling and analysis of hybrid dynamical systems is becoming more and more important as such systems are now widely used to reason about complex physical systems. Ying Zhang and Alan Mackworth developed a semantic model for dynamic systems, called Constraint Nets (CN) [1]. CN introduces an abstraction and unitary framework to model hybrid systems. Furthermore, specification and verification methods were introduced for deterministic system.

Traditional approaches to real-time hybrid systems usually define behaviors purely in terms of determinism or sometimes non-determinism. The CN framework was develop to model and verify deterministic systems, with the capability to model non-determinism. However, real-time dynamical systems very often behave probabilistically and thus exhibit (structured) uncertainty. It is therefore important to be able to model and analyze real-time probabilistic systems. Hence, a formal framework to model systems with unpredictable behaviors is essential.

We extend the work previously done on Constraint Nets by developing a new framework that we call “Probabilistic Constraint Nets” (PCN). The PCN framework allows for the modeling and simulation of any dynamical system, whether it is deterministic, non-deterministic or probabilistic. We introduce formal syntax and semantics for the framework that ensure the correctness of the models. We also provide a graphical representation that simplifies the task of modeling complex systems. Moreover, we show that our framework is a generalization of many commonly used frameworks like Bayesian Networks and Markov Decision Processes (MDP) which allows the user to take advantage of a unified framework encompassing most popular modeling paradigms. Furthermore, we provide, for a subclass of PCN models called synchfin-PCN, algorithms for control synthesis. With such algorithms, a designer can automatically construct an optimal controller for his system, hence greatly facilitating his task.

Current work include the formulation of a specification language and verification algorithms for PCN models while upcoming work will tackle the task of coupling the PCN framework with learning techniques that would allow a designer to build a controller capable of adjusting its systems parameters as it interacts with the environment in a real time fashion.

[1] Ying Zhang and Alan K. Mackworth. Constraint nets: a semantic model for hybrid systems. Journal of Theoretical Computer Science, 138(1): 211-239, 1995. Special issue on hybrid systems.

Dynamic Manipulation of Perception of Contact with Leading Automobiles Through Control of Optical Looming Cues

Zhonghai Li, Ph.D. Student Candidate
Ergonomics in Teleoperation and Control (ETC) Laboratory
University of Toronto, Toronto, M5S 3G8

Rear-end collisions are (unfortunately) a very common type of traffic event, which occur when a following vehicle (FV) strikes the rear of a lead vehicle (LV). Whenever an observer is on a collision course with another object, the opposing object generates a characteristic spatiotemporal expansion on the observer's retina. The part of the approaching surface with which one will eventually make contact (if no steps are taken to prevent the collision) expands symmetrically in terms of its retinal image size. The resultant effect is called optical looming, and is an important cue for the perception of contact with objects [1].

In terms of manipulation of optical looming cues, five important earlier studies [2] [3] [4] [5] [6] have been conducted. In all of those it was found that direct real-time manipulation of the looming cues of a target (i.e. object expansion or contraction during approach) can influence the perception of contact with the target (e.g. approaching balls, barriers, etc.). The result is that subjects (animals or humans) respond (e.g. by grasping, braking, etc.) earlier to an expanded target or later to a contracting target, relative to a constant size target.

Experiments are being developed to test this concept for automobile driving in a driving simulator. In our 2x2 experiment, the first factor to be manipulated is size of the lead vehicle (LV), such as a leading car and truck. The second factor is *dynamic* manipulation of the size of whole lead vehicle during the experiment. If a significant effect is found with respect to perception of time to contact with LV, a further experiment will be conducted involving manipulation of vehicle size, illumination, driving scenario and LV taillight geometry. Our ultimate aim is to explore the possibility of manipulating perception of contact with leading automobiles, potentially through control of the separation of LV taillights, one of the primary cues used to regulate closure during vision impaired driving (e.g. fog, night-time).

If our hypothesis proves correct, introduction of the concept of dynamic taillights to actual vehicles on the road could cause a significant number of following drivers to brake sooner, and so reduce the frequency of rear-end collisions. Such a manipulation would presumably happen only when the lead vehicle brakes suddenly and/or forcefully. Equally important is our hypothesis that such a mechanism will be effective only if these manipulations are imperceptible for the following driver, in order to exclude the possibility of adjustment of subjective risk criteria and eventual entrainment. Clearly, a decrease of even a few tens of milliseconds in braking times could have a significant impact on overall traffic safety, when aggregated over the extremely large number of FV braking incidents.

[1] Gibson, J.J. (1979). The ecological approach to visual perception. Boston: Houghton Mifflin.

[2] Savelsbergh, G. J. P., Whiting, H. T. A., & Bootsma, R. J. (1991). "Grasping" tau!. *Journal of Experimental Psychology: Human Perception and Performanc*e, 17, 315- 322.

[3] Savelsbergh, G. J. P., Whiting, H. T. A., Pijpers, J.R. and van Santvoord, A.A.M. (1993). The visual guidance of catching. *Experimental Brain Research* 93, 148-156. [4] Sun, H.J. and Frost, B.J. (1998b). Visual control of target-directed locomotion in a virtual environment. IRIS/Precarn Conference.

[5] Wann, J. & Rushton, S. (1995). Grasping the impossible: stereoscopic virtual balls. *Studies in perception and action.* B.G.bardy, R.J.Bootsma & Y. Guiard (Eds). Lawrence Erlbaum Association, Inc.

[6] Sun H-J, Carey D. P. & Goodale M.A. (1992). A Mammalian model of optic-flow utilization in the control of locomotion, *Experimental Brain Researc*h, 91: 171-175.

Simultaneous Localization and Mapping with Multiple Mobile Robots

Ruslan Salakhutdinov
Department of Computer Science
University of Toronto
6 King's College Rd, M5S 3G4, Canada
rsalakhu@cs.toronto.edu

Sam Roweis
Department of Computer Science
University of Toronto
6 King's College Rd, M5S 3G4, Canada
roweis@cs.toronto.edu

By identifying each state in a Hidden Markov Model (HMM) with a small some spatial region of a continuous space, it is possible to naturally define neighbouring states as those which correspond to connected regions in the underlying space. The transition matrix of the HMMcan then be constrained to allow transitions only between neighbours; this means that all valid state sequences correspond to connected paths in the continuous space. We apply this constrained HMM architecture [2] to the problem of simultaneous localization and mapping (SLAM) from sensor logs of mobile robots navigating in unknown environments. For a single robot, or multiple non-interacting robots, the learning and inference algorithms are identical to those for standard HMMs trained on multiple observation sequences, except that the transition matrix is fixed by the spatial topology and is not up-dated during learning. The more interesting case is that of multiple robots which explore the environment simultaneously and interact with each other, for example through proximity sensors. In this case, exact learning requires inferring the joint state of all robots, and quickly becomes exponentially expensive because the effective state space of the HMM is the product of the state spaces of the individual robots. We develop an approximate method for inference in this case using belief propagation [1] and apply it to the SLAM problem with multiple interacting robots, showing that with the same amount of data, approximate learning with the interaction signals outperforms exact learning ignoring interactions.

References

[1] Brendan J. Frey and David J. C. MacKay. A revolution: Belief propagation in graphs with cycles. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10. The MIT Press, 1998.

[2] S. T. Roweis. Constrained hidden markov models. In *Advances in neural information processing systems*, volume 12, pages 782–788, Cambridge, MA, 1999. MIT Press.

Optimizing the Secure Evaluation of Queries Based on XML DTD Graph

Hui Wang, Ph.D Degree Student
Database Lab
University of British Columbia, V6T1Z4

Because of the increasing amounts of real applications that sensitive data is stored in XML documents, there are a few problems arising related to XML access control to protect those sensitive data. One of the problems is to ensure that user queries only use and return the XML data he is allowed to access. The naive method is to check on each element in query to ensure its accessibility, which is much expensive.

Our research focuses on the simple, but useful multi-level access control model, where a security level can be either specified at an XML element, or inherited from its parent. For this model, we devise efficient algorithms to eliminate expensive recursion functions on as many query elements as possible based on the knowledge from XML DTD and still ensure the correctness of the accessibility on the elements.

On Visual Maps and Their Automatic Construction

Robert Sim

An important prerequisite for an autonomous robot is the ability to explore its environment and construct a map which is useful for planning and inference. Autonomy is desirable in a variety of application domains, including search and rescue, hazardous environment inspection and sea and space exploration. Our work has two main contributions. First, we develop the concept of the visual map, a representation of the visual structure of the environment. Such maps are useful for a robot equipped with a monocular camera to meet the planning and inference prerequisites for autonomy. Second, we present a method for automatically exploring an environment and constructing a visual map in the face of uncertainty. These approaches are presented in the context of learning which visual features of the world are best suited for robot pose estimation, and subsequently learning which exploratory trajectories are best suited to acquiring such features while minimizing uncertainty.

The core concept of our work is that of the visual map. We demonstrate how image-domain features can be selected, modeled and evaluated for their utility for pose estimation. Feature selection entails employing a model of visual attention to locate salient features, followed by a tracking procedure that collects feature observations across a variety of robot poses. Given these observations, the visual features are modeled as a function of pose using an interpolation framework. The resulting models are generative in nature, allowing the robot to predict feature behaviour from novel views. The resulting models are evaluated for their reliability, and can be employed for inference tasks such as pose estimation and scene reconstruction. We present experiments demonstrating the feature learning process and their reliability for pose inference.

The second part of our work addresses the exploration and construction requirement for autonomy. First, we demonstrate that visual maps are self-organizing in nature, requiring only an initial odometric input to establish the transformation between the image and pose domains. Second, we demonstrate how visual maps can be constructed reliably in the face of uncertainty by selecting an appropriate exploration strategy. We present a variety of such strategies and discuss their efficiency and reliability. These approaches are again validated using experimental results.

Curve Synthesis From Learned Refinement Models

Saul Simhon and Gregory Dudek
Centre for Intelligent Machines
McGill University

We develop a method for augmenting the shapes of noisy and coarse curves using details from a-priori data.

While our primary focus is on sketch understanding and refinements, our system can be used for applications such as the de-blurring of edges, robot motion control and animation. Noisy curves, which can be extracted from an arbitrary source (such as a set of edges from an edge detector, hand-drawn curve strokes from a drawing or the path from a robot path planner) are augmented to look more like examples selected from ensembles of desired curve motions. These examples can encompass various types of constraints, primarily encoding the desired shape patterns. Application specific constraints can also be included such as curve color/thickness or robot axis and facing direction.

The shape is represented by a sequence of tangent angles. This sequence is considered to be generated by a stochastic process exhibiting the Markov assumption (i.e. only local consistency is enforced). Additionally, each of these stylized curves are coupled with a 'control' curve that identifies the transformation between the inputs expected and desired output. For example: a horizontal line is associated with a ziz-zag pattern, a vertical line is associated with wave pattern, a sharp turn is associated with a smoothed arc-like turn. The examples are used to train a Hidden Markov Model where the observation states play the role of sample points from the control curves and the hidden states play the role of sample points from the desired stylized curves. Using the HMM in combination with multi-scale methods and Gaussian Mixture Models, we compute a probability distribution for successive curve segments as a function of their context within the stylized curves and the control curve that guides them. Given an input curve, a novel stylized curve is generated by choosing some mixture of the training set biases that produces the maximum likelihood estimate.

Given that we may have several training sets depicting various styles and motions, we wish to automatically detect what sets to use for the input curves. Our approach consists of using a two-level Hierarchical HMMs. The lower level in the hierarchy consists of the HMMs trained using the curve sets. The Higher level of the hierarchy moderates the recognition of what sets should be used by specifying the conditional probability of having one type of curve after another, or one type below or above another. Given a set of input curves, we first classify each curve as belonging to one of the sets based on its shape and context (i.e. the higher level HMM), then we generate the refined curves by applying the appropriate HMM to each.

Automated Perimeter Estimation for Procedural Texturing

Eric Bourque
Centre for Intelligent Machines
McGill University
Montreal, QC H3A 2A7

This poster describes an approach to automatically select a procedural texture and associated parameters to replicate a given sample texture from an input image or sketch. In general, selecting a procedural texture from a library or setting the parameters for one can be complex and time consuming. We propose a strategy for automating this process. While, superficially, the problem appears intractable for both humans and computational systems, its natural characteristics make a computational solution feasible. We present an algorithm and experimental results showing the feasibility of our approach which we call image-based procedural texturing.

Range Synthesis for Mobile Robot Environment Modeling

Luz A. Torres-Mendez and Gregory Dudek
Center for Intelligent Machines
McGill University, Montreal, QC, H3A 2A7

In mobile robotics, the use of range data for navigation and mapping has become a key methodology. However, the acquisition of complete range maps (i.e. volume scans) from a single view point remains prohibitive for many real systems. Stereo cameras can produce volumetric scans that are economical, but they often require calibration or produce range maps that are either sparse or of limited resolution. Volumetric laser scanners tend to be costly and physically demanding or slow. A particular common simplifying assumption is to represent 3D structure as a 2D "slice" through the world. However, in practice this is not sufficient to capture structures of interest.

In this work, we present a novel statistical learning method for computing range data as an initial solution to the environment modeling problem in the context of mobile robotics. Unlike other methods that are based on a set of geometric primitives, our method computes dense range maps of locations in the environment from a combination of a intensity image and a limited amount of observed range data. This should allow a robotic system to rapidly collect a small amount of range data, for example, from as little as a few scans of a laser line-striping device across the environment and a video image, and then infer the rest of the range map it does not capture directly.

Our methodology is to statistically learn the relationship between the observed range data and the variations in the intensity image and use this to compute the unknown range data. This can be regarded as a form of shape-from-shading based on statistical learning, although traditional shape-from-shading is quite different from this approach in its technical details. In our approach, we approximate the composite of range and intensity at each point as a Markov process. Unknown range data is then inferred by using the statistics of the observed range data to determine the behavior of the Markov process. The presence of intensity where range data is being inferred is crucial since intensity data provides knowledge of surface smoothness and variations in depth. Our approach learns that knowledge from the observed data, without having to hypothesize constraints that might be inapplicable to a particular environment.

We base our range estimation process on the assumption that the pixels constituting both the range and intensity images acquired in an environment, can be regarded as the results of pseudo-random processes, but that these random processes exhibit useful structure. In particular, we exploit the assumption that range and intensity images are correlated, albeit in potentially complicated ways. Secondly, we assume that the variations of pixels in the range and intensity images are related to the values elsewhere in the image(s) and that these variations can be efficiently captured by the neighborhood system of a Markov Random Field. Both these assumptions have been considered before [1,2], but they have never been exploited in tandem.

Our approach was tested using data from a real environment given initial data in various configurations. The obtained results demonstrate that this is a viable option to facilitate the environment modeling.

[1] A. Efros and T.K. Leung. “Texture synthesis by non-parametric sampling”. In ICCV (2), pages 1033-1038, September 1999.

[2] S. Geman and D. Geman. “Stochastic relaxation, gibbs distributions, and the bayesian restoration of images.” IEEE Transactions on Pattern Analysis and Machine Intelligence, 6:721-741, 1984.

An Inexpensive 122 Million Pixel Scan Camera

Shuzhen Wang
University of British Columbia
C/O MAGIC, FSC 3940, 2424 Main Mall B
Vancouver, BC V6T 1Z4

Phone: 604-822-4231(office)
604-822-2218(lab)
Email: szwang@cs.ubc.ca

In this poster, we introduce the design of a low-cost, very high-resolution scan camera. This camera can take black/white, near-infrared, and color images with the resolution of 122 million pixels. Firstly, the motivation of this project is given, then we describe the hardware setup and calibration software. Finally, we give some samples images from the scan camera.

Cloth Motion Capture

David Pritchard
Department of Computer Science
University of British Columbia
Name of supervisor: Wolfgang Heidrich

Recent years have seen an increased interest in motion capture systems. Current systems, however, are limited to only a few degrees of freedom, so that effectively only the motion of linked rigid bodies can be acquired. We present a system for the capture of deformable surfaces, most notably moving cloth, including both geometry and parameterisation. We recover geometry using stereo correspondence, and use the Scale Invariant Feature Transform (SIFT) to identify an arbitrary pattern printed on the cloth, even in the presence of fast motion. We describe a novel seed-and-grow approach to adapt the SIFT algorithm to deformable geometry. Finally, we interpolate feature points to parameterise the complete geometry.

Automatic Acquisition of Motion Trajectories: Tracking Hockey Players

Kenji Okuma, Masters Degree Student
The Laboratory of Computational Intelligence
The University of British Columbia
Vancouver, British Columbia, V6T 1Z4

Computer systems that have the capability of analyzing complex and dynamic scenes play an essential role in video annotation. Scenes can be complex in such a way that there are many cluttered objects with different colors, shapes and sizes, and can be dynamic with multiple interacting moving objects and a constantly changing background. In reality, there are many scenes that are complex, dynamic, and challenging enough for computers to describe. These scenes include games of sports, air traffic, car traffic, street intersections, and cloud transformations.

Our research is about the challenge of inventing a descriptive computer system that analyzes scenes of hockey games where multiple moving players interact with each other on a constantly moving background due to camera motions. Ultimately, such a computer system should be able to acquire reliable data by extracting the players' motion as their trajectories, querying them by analyzing the descriptive information of data, and predict the motions of some hockey players based on the result of the query.

Among these three major aspects of the system, we primarily focus on visual information of the scenes, that is, how to automatically acquire motion trajectories of hockey players from video. More accurately, we automatically analyze the hockey scenes by estimating the panning, tilting and zooming parameters of the broadcast cameras, tracking hockey players in those scenes, and constructing a visual description of the data by displaying trajectories of those players. Many technical problems in vision such as fast and unpredictable players' motions and rapid camera motions make our challenge worthwhile to be tackled. To the best of our knowledge, there have not been any automatic video annotation systems for hockey developed in the past. Although there are many obstacles to overcome, our efforts and accomplishments would hopefully establish the infrastructure of the automatic hockey annotation system and become a milestone for research in automatic video annotation in this domain.

AquaSim

Arlene Ripsman, Ph.D. Student
Department of Computer Science and Centre for Vision Research
York University, Toronto, M3J 1P3

An area of research that is currently in its infancy is three-dimensional path planning with robots that autonomously explore their environment, using two modalities. For example, robots that can both walk and fly or walk and swim. Since we are dealing with a three-dimensional world with robots that possess two options of locomotion, more complex path planning algorithms, than those used in a simple two-dimensional world, must be employed.

A current project looking at a robot that employs two modalities is the AQUA project. This research involves creating an underwater version of the RHex robot [2] that can explore a lake environment both by walking on the bottom of the lake and by swimming.

A significant challenge associated with this research is the difficulty in testing different path planning algorithms in the field. In order to properly test path planning algorithms for the aquatic RHex robot, a well designed simulator is required. This research focuses on the development of such a simulator. This simulator, referred to as AquaSim, will be used to test three-dimensional path planning algorithms for the AQUA project. AquaSim is currently under development to run within the VE environment [1]. The VE environment permits the simulation to be run on a variety of different hardware platforms, including IVY, York University's six-sided fully immersive display. AquaSim uses a dynamic model of the aquatic RHex robot, as well as a dynamic model of a lake environment to be explored. Three-dimensional path planning algorithms are being developed and run to simulate the robot's exploration of its environment. The results of each algorithm are observed and evaluated.

[1] M. Robinson, "The VE User Manual", http://www.cs.yorku.ca/~ivy/ve.

[2] U. Saranli, M. Buehler, D.E. Koditschek, "RHex: A Simple and Highly Mobile Hexapod Robot", International Journal of Robotics Research, 20(7):616-631, July, 2001.

A Software Framework for Developing High Quality Control Systems for Autonomous Robots

Darrell M. Lahey, James J. Little
Department of Computer Science, University of British Columbia
{lahey,little}@cs.ubc.ca

Quality is an important issue in developing software systems. The best systems support all the functionality required by their users, while still maintaining high quality. This idea applies equally to robot control systems. For users of robot control systems, the most relevant measures of quality include correctness, usability, efficiency, reliability, integrity, adaptability, and robustness.

For developers, the most relevant measures include extendability, maintainability, flexibility, portability, reusability, testability, understandability, and compatibility. Strong cohesion and loose coupling help to improve system quality. Modern software engineering practice, which includes object-oriented programming, helps developers to construct high quality systems. It also allows developers to redesign existing systems such that they have higher quality. We have derived a software framework that allows developers to create high quality control systems for autonomous robots. Systems developed using this framework are behavior-based, and each system consists of a network of components. This network is essentially a directed graph, where the vertices correspond to the components, and the edges correspond to connections between components. All vertices are types of objects. They are equivalent to objects in object-oriented programming, and they store data that is required for deliberative tasks. Skills are special objects that contain collections of behaviors. Behaviors repeatedly take input from one set of objects, and then modify another set of objects, by calling methods on objects. With this processing, behaviors facilitate the development of reactive systems. Skills impose input and output restrictions on their behaviors. The input for the behaviors comes from specified objects, and these input constraints are represented by directed edges from objects to skills. Behaviors generate output by calling the methods on specified objects, and these output constraints are represented by directed edges from skills to objects. Task sequencers are special behaviors that can set tasks for other skills to perform. The ability of a task-setting skill to control a task-performing skill is represented by a directed edge from the task-setting skill to the task-performing skill. Skills, with task sequencers, simplify the development of deliberative systems, and they promote strong cohesion and loose coupling. Specifically, they allow behaviors with similar functionality to be managed as single, cohesive, units. In addition, task sequencers control skills, instead of controlling individual behaviors. Filters can be attached to objects such that they intercept method calls and modify parameter values. They facilitate coordination and robustness. Modules allow sub-networks to be specified as special skills. They facilitate reusability within a system.

We implemented our framework in Java because we have determined that it has the best built-in support for our framework. By providing its Remote Method Invocation (RMI) and Reflection classes, Java allows developers to write code, for behaviors and objects, that does not contain any networking details. Because our framework is Java-based, developers must write behaviors and objects using Java source code. However, through the Java Native Interface, this source code can invoke routines that have been written in other programming languages. In addition, users can write code that cooperates with other robot control systems. Our implementation provides a configuration manager, which allows developers to specify, and execute, robot control systems.

Automatic Recognition of Hand Raising Gestures and Voice Requests for a Remote Learning Application

B. Kapralos(1,2) and M. Jenkin(1,2)
1. Deptartment of Computer Science, York University, Toronto, Ontario, M3J 1P3
2. Centre for Vision Research, York University. Toronto, Ontario. M3J 1P3

A central issue in developing synchronous distance learning technology, especially with students and instructors at a number of different locations, is enabling the remote classes and the instructor to interact with each other. There are really two parts to this problem: How to present the instructor to the remote classrooms and how to present the remote classrooms to the instructor. The first part of this problem is perhaps the easiest to solve, as there is only one person (the instructor) who must be attended to. Attending to students in the remote classrooms is more difficult. Issues such as "How does a student `raise his/her hand' for attention?" or "How can the instructor `select' one student to converse with, and how to attend to the student once (s)he has been selected?" are complex problems which must be addressed if the classes and instructors are to interact in an effective manner. Providing human facilitators at each site is not cost effective and the option of physically wiring each seat with buttons for students to indicate that they have a question would require significant modifications to existing classroom spaces. An alternative would be to deploy a sensor system within the classroom that enables student interaction with the instructor and other classes. But how should the sensor attend to the person who wishes to ask a question? From a practical point of view, how should a sensor be constructed that has a wide enough field of view so that it can capture the entire class at once and also be able to attend to the person who wants to speak or ask a question? In addition, once a speaker has been selected, how should the sensor continue to track, localize and focus on the selected speaker?

This research project is investigating issues related to the development of a distance learning system that permits a remote classroom to interact with the instructor. This includes issues related to attending, in both the audio and visual domain, to individual students, finding students who wish to speak, permitting the instructor to view the entire remote class and to attend to audio and visual events within it. To this end, a novel sensor that combines directional audio and a panoramic sensor to locate students in the classroom who wish to interact with the instructor has been developed. Students who wish to interact with the remote instructor may signal their intent via voice (e.g. by speaking aloud) or by using hand raising gestures. Hand raising gestures are detected using a combination of color and motion cues over a sequence of omni directional images while sound a steerable microphone array allows for the detection of audio signals. The system identifies "attention seeking" actions in the audio and video domain, and then presents potential speakers to the remote instructor. The instructor can then select, via a touch-screen based user interface, which one of the potential speakers (including speakers who have not sought attention overtly) and the sensor will then attend to that speaker in both the audio and visual domains. A high resolution view of the speaker and a beamformed audio signal can then be presented to the remote instructor.

CinDeR - Collision and Interference Detection in Real-time Using Graphics hardware

Dave Knott, Graduate Student
Department of Computer Science
University of British Columbia
Vancouver, British Columbia

A vital task in almost all forms of computer animation or physical simulation is the act of *collision detection*[2]. When polyhedral objects are being animated, it is critical to determine if and when they are about to, or have already, come into contact with each other. Example problem domains where collision detection is almost ubiquitous include rigid and deformable body simulation, computer games, virtual reality, surgical simulation, robotics, path planning, and computer-aided design and manufacturing (CAD/CAM). Collision detection is also one of the most computationally demanding tasks in each of these domains and therefore one of the most common bottlenecks in the simulation pipeline.

In recent years, the computational power of graphics hardware has made enormous leaps, not only in speed but also in functionality. A typical graphics processor now contains more logical transistors than the CPU of the computer that it resides in.

It is reasonable, then, to ask whether the computation involved with collision detection can be offloaded from the computer's primary processor and memory, and be performed instead on the graphics hardware. To do so frees up CPU power for other tasks and enables the application to make use of a computational resource that might otherwise be underutilized.

We have developed an image-space method for detecting interference between solid polyhedral bodies[1]. The algorithm makes use of virtual ray casting to determine which portions of the edges of the polyhedrons in question lie within volumes enclosed by other polyhedrons. The ray-casting is performed by rendering the objects using 3D graphics acceleration hardware and using a variety of depth buffer, stencil buffer, and colour buffer operations.

The technique exhibits a number of features which, to our knowledge, no other interference detection algorithm has successfully combined:

- Convex and non-convex geometry with hollow regions can be handled.
- Large numbers of objects can be handled.
- Intersection tests are performed on the geometry itself, not on an approximation to the surface.
- No special data structures are required.
- No preprocessing of models is required.
- The algorithm's expected asymptotic running time is linear in both the number of objects being tested and the number of polygons comprising the objects.
- Processing is done with the aid of commodity-level graphics hardware.

[1] D. Knott, and D. K. Pai. CInDeR: Collision and Interference Detection in Real-time using graphics hardware. In *Proceedings of Graphics Interface 2003*

[2] M. Lin and S. Gottschalk. Collision Detection between Geometric Models: A Survey. In *Proceedings of IMA Conference on Mathematics of Surfaces*, 1998

Substituting Touch for Vision

Sam Bromley, School of Engineering
University of Guelph

Professor John Zelek, IRIS Emerging Opportunities Fund Project

Passive stereo machine vision can provide robust spatial information about an environment provided that the underlying computational algorithms are sufficiently sophisticated. Significant difficulties arise when the spatial information is required for real-time tasks such as navigation, as the processing time available is then quite limited. The goal of this research is to develop the mathematical models, computational algorithms, and technological devices required to both generate real-time, robust depth information about the environment, and to convey this information to a mobile robot platform, or more importantly, to a blind person via a tactile feedback device. In both cases, the resultant information transferred will be band limited, in the sense that not all of the scene information is required for successful navigation, and moreover, would lead to information overload for the human or robot if conveyed in its entirety. Ideas borrowed from probabilistic particle filtering, coupled with more classic stereo vision methods, should permit a novel stereo machine vision approach with unprecedented robustness for navigation applications. The goal is to develop a reduced data set of important information sufficient for successful navigation, along with the machine vision algorithms that can generate this data for arbitrary scenes, at video rates, on commodity hardware.

Light-Weight Semantic Matching for Web-Service Discovery

Yiqiao Wang,
Masters Degree Student
221 Athabasca Hall,
University of Alberta,
Edmonton Alberta, T6G 2E8

Faced with decreasing time-to-market and increasing requirement volatility, software-development processes are increasingly relying on reuse of existing software. Furthermore, the World Wide Web is increasingly being adopted as the medium of collaboration with partners and as a means of delivering information and services to consumers; thus, web-based applications constitute a substantial percentage - if not the majority - of the currently developed applications. The web-services stack of standards is designed to support the reuse and interoperation of software components on the web. It consists of a set of related specifications, defining how reusable components should be specified (through the Web-Service Description Language - WSDL [1]), how they should be advertised so that they can be discovered and reused (through the Universal Description, Discovery, and Integration API - UDDI [2]), and how they should be invoked at run time (through the Simple Object Access Protocol API - SOAP).

A critical step in the process of reusing existing WSDL-specified components is the discovery of potentially relevant components. UDDI, the standard API for publishing web-services specifications, currently provides only a simple browsing-by-business-category mechanism for exploring and discovering published services. This service-discovery method through category selection method is clearly insufficient. It is quite informal and relies, to a great extent, on the shared common understanding of publishers and consumers. It is the responsibility of the provider developer to publish the services in the appropriate UDDI category. The consumer developer must, in turn, browse the "right" category to discover the potentially relevant services. More importantly, these methods do not provide any support for selecting among competing alternative services that could potentially be reused.

In this poster, we introduce WSDLMatcher, a signature-matching algorithm for discovery and selection of WSDL services. The underlying intuition of this work is that an alternative means of querying UDDI servers is "by example", i.e., by providing a potentially partial specification of the desired service to retrieve similar published services. Using our algorithm, the consumer developer may define various aspects of the desired service, such as namespaces of input/output information for example, and the algorithm will return a set of candidate services with an estimate of their similarity to the provided example, thus enabling a more informed selection by the developer. This query-by-example process for WSDL-service discovery and selection involves the following steps:

1. Receive a (potentially partial) WSDL specification of a desired service (example service).

2. Access WSDL specifications from UDDI registry (candidate web services).

3. WSDLMatcher matches example WSDL service against all candidate WSDL services based on their interface and returns estimated similarity scores between the two services. Further steps include applying traditional techniques from IR combined with WordNet [3], an on-line lexical database for the English language, to further understand the semantics of web services described. Semantic web efforts such as LARKS [4] propose a full-fledged ontology for defining the domain-specific semantics of web services, and this definition process is extremely costly. We will be investigating how effective light-weight natural-language based semantics combined with signature matching would be at a much lower cost.

On The Design Of Caching Schemes For Mobile Devices

Mildred Ambe, Ioanis Nikolaidis, Eleni Stroulia
Computing Science Department
University of Alberta
Edmonton, AB T6G 2E8
mildred@cs.ualberta.ca,yannis@cs.ualberta.ca,
stroulia@cs.ualberta.ca

Developing software for wired network devices is facilitated by the processing power and memory capabilities available today. In contrast, a whole new set of challenges arises when developing software for wireless devices: small screen size, limited memory and low processing power. Numerous applications on small devices today require data from the Internet, and one major concern is the unreliability of wireless connections. The latter is due to possible disconnections, high bit error rate, and low bandwidth [2]. A lot of research has been reported on the performance of Web Caching for wired Internet access. However, this problem is substantially different in the context of wireless devices: the nature of the tasks performed on mobile devices is different, and also the interaction style is different. The objective of this project is to design a general test-bed for experimenting with and analyzing different caching schemes that can be used with mobile device and the Java 2 Micro Edition platform (J2ME).

To establish a control baseline for our experimentation framework, we first developed a simple J2ME Web browser and a test web access log, consisting of a set of web sites with a variety of HTML documents. All the functionalities of this first simple browser were implemented on the mobile device, including user-interaction support, URL access, web-data retrieval, parsing of HTML content, and display of the parsed content. We then used the browser to access and display the HTML-document sequence in our test suite. Not surprisingly, the browser proved to be quite inefficient, because of the limited device memory and slow connections to the WWW servers.

Since this original experiment, we have redesigned and developed a second browser as a client-server application. This application enables all computationally expensive and communication intensive activities, such as data fetching and parsing, to be handled on the server that acts as a proxy between the mobile client and the web servers it accesses. This new application is our experimental test-bed.

In principle, caching of visited documents and prefetching documents that are expected to be of interest are the most common approaches for coping with low bandwidth. Currently, we have focused our research on caching schemes. *Caching* allows for the storage of accessed web pages in a local cache structure, so that when they are requested in the future they do not have to be fetched again. When the local cache is full, an eviction policy decides which of the existing documents will be deleted to make space for the new document to be cached. In this poster, we discuss our experiments with several caching scenarios, which can be classified along two dimensions:

(a) Where is the information cached (client browser or proxy server) and

(b) What is the basis of the eviction policy (least recent first, least frequent first, based on the importance of the document in the context of the web-site structure, or based on user-behavior patterns).

Good caching schemes will demonstrate high hit rates and byte-hit rates from the cache. The objective of this experiment is to distinguish the schemes that maximize efficiency by reducing the number of accesses made to WWW servers.

References

[1] Z. Jiang, L. Kleinrock, "Web prefetching in a Mobile Environment", U. of California at L.A.

[2] P. Loganathan, "The Challenges and Requirements of Mobile Computing", ISE II, Imperial College of Science, Technology and Medicine.

[3] Java Sun Site (J2ME) – http://java.sun.com/j2me

Three Dimensional Ultrasound And Stereotactic Mammography Guided Biopsy: A Dual Modality System

Kathleen JM Surry, Doctoral Candidate
Collaborating Authors: A Fenster, DB Downey
Imaging Research Laboratory, Robarts Research Institute
London, Ontario, N6A 5K8

Advances in image guidance for breast biopsy have allowed clinicians to choose a minimally invasive option for evaluating suspicious masses, instead of surgical intervention. The gold standard in image guidance for needle biopsy is stereotactic mammography (SM). Alternatively, ultrasound (US) imaging may also be used to guide the needle to a target. This real-time, two-dimensional (2D) modality can be interpreted in terms of the relative three-dimensional (3D) geometry of the breast and the needle trajectory. This technique requires an expert operator for both safety and accuracy, because the trajectory of the needle is unconstrained and the tip of the needle may be misinterpreted in the US image. The purpose of this work was to develop a three-dimensional ultrasound-guided biopsy system to supplement stereotactic mammography with near real-time 3D and real-time 2D ultrasound imaging. Features from SM and US guided biopsy have been combined, including breast stabilisation, a confined needle trajectory and dual modality imaging. Our goal is to be able to biopsy a larger percentage of suspicious masses using ultrasound, by clarifying ambiguous structures with mammographic imaging. We have evaluated our prototype system using breast phantoms, in terms of its accuracy with ultrasound guided biopsy [1]. We have also registered stereotactic mammography images to the 3D US images. We have shown that our ultrasound guided biopsy system was capable of placing the needle tip with 0.85 mm accuracy at a target identified in the 3D image. We also identified that we could successfully biopsy artificial lesions that were 3.2 mm in diameter, with a 96% success rate. Metal beads in known relative positions allowed us to determine the geometry of the system. The target registration error (TRE) was found to be 4.8 mm. This error is dominated by a large positioning error in the vertical direction (perpendicular to the film surface). The horizontal components (x and y) of the TRE were 0.36 mm and 0.26 mm. The vertical (z) component was found to be 4.8 mm. Using an ultrasound guided biopsy apparatus, we are able to accurately target artificial lesions in breast phantoms. As an adjunct to stereotactic mammography, we propose that this system could provide more complete imaging information for target identification and real-time monitoring of needle insertion, as well as providing a means for rapid confirmation of biopsy success with three-dimensional ultrasound.

[1] K.J.M. Surry, W.L. Smith, L.J. Campbell, G.R. Mills, D.B. Downey, A. Fenster. "The development and evaluation of a three-dimensional ultrasound-guided breast biopsy apparatus". Medical Image Analysis, 6:301-312, 2002.

Distributed Embedded Video Processing Hardware Solution for Local Positioning System and Other Computer Vision Research Application

Adnan Mir Ali, Zion Kwok, Florian Vogt, Baosheng Wang
Human Communication Technologies (HCT) Research Laboratory
Department of Electrical and Computer Engineering
University of British Columbia, Vancouver, BC, V6T 1Z4

Project: Parallel Distributed Camera Arrays
Supervisor: Sidney Fels

People and object tracking has been the focus of much research and product development often referred to as automatic identification and data capture (AIDC) technology. Different solutions and techniques for tracking are implemented in existing systems. Each approach has solved the technical challenges by carefully considering the application domain [1]. For example, typical tracking for motion capture has used either optical markers or magnetic field sensing (Ascension's Reactor and Polhemus Fastrak systems). Our Local Positioning System (LPS) is designed to track people-sized objects in large, indoor spaces.

In this project, our goal is to design a research platform for computer vision applications, as part of the IRIS distributed camera array project. The system consists of a number of distributed, embedded video processing units that implement computer vision algorithms. Each processing unit contains a video camera, processing hardware, as well as a network interface used to communicate

data and events to other units or a host node. The overall design targets the requirements of the Local Positioning System (LPS) application while maintaining flexibility for its role as a computer vision research platform. The LPS is a scalable, accurate, and low-latency-tailored vision based system for tracking the 3-D position of multiple objects using IR tags in a small area such as a room or exhibition hall.

Two hardware systems were designed to support the LPS and research platform. The first camera system uses a microcontroller that reads the image from the camera, executes a video processing algorithm, and communicates the results to another node through an Ethernet. The second camera system adds the power of an FPGA. The FPGA reads the image from the camera, performs initial video processing, and passes the results to the microcontroller for additional processing and network communication. Although the first system is sufficient to implement the full functionalityof the Local Positioning System, it is limited by speed constraints imposed by the I/O interface. The second system will therefore emerge as the chief platform for both the LPS application and the computer vision research.

The second system excels in its flexibility provided by the microcontroller software in combination with the processing power of the FPGA. Currently, we are using a 400 MHz ARM microcontroller accompanied by large memory (64MB), typically scarce among embedded systems. Besides simplifying firmware development, this processor also allows the use of common operating systems such as Linux, WinCE or ECOS, which offer a familiar environment as well as code libraries. The micro-controller can adjust system parameters, track moving objects, and perform complex processing that is more efficient in software. A distributed system can scale for high-speed, high-resolution applications, low-speed, low-resolution applications, and applications with various other requirements. A cost-effective but advanced FPGA was chosen for its ability to implement high speed video processing algorithms. The available power is much higher than required for the LPS, however, the additional speed, memory and logic provides flexibility for other applications that are part of the PDCA project.

A prototype of the first system consists of an Intrinsyc Cerf ARM CPU board, a black and white OmniVision camera connected to general I/O pins, and a 100BaseT Ethernet connection. It is capable of capturing images from the camera and processing them for the LPS application. A prototype for the second system requires an additional circuit board that supports an Altera ACEX FPGA that we have built. The current camera we support runs at 50 frames per second with 384x288 resolution. Our architecture is meant to scale as different cameras are needed for more advanced applications. Thus, our platform provides for future improvements involving cameras that are faster and/or offer higher resolution and optimization configurations for size and cost.

[1] Jeffrey Hightower and Gaetano Borriello (2001) Location Systems for Ubiquitous Computing, Computer Journal, IEEE Computer Society Press, 34(8), pages 57-66.

Simulation of Distributed Parallel Camera Arrays For Computer Vision Applications

Adnan Mir Ali and Florian Vogt
Human Communication Technologies (HCT) Research Laboratory
Department of Electrical and Computer Engineering
University of British Columbia, Vancouver, BC, V6T 1Z4

Project: Parallel Distributed Camera Arrays
Supervisor: Sidney Fels

The development and implementation of distributed parallel camera arrays presents unique problems to the developer. Distributed camera arrays generally need to be calibrated, statically or dynamically, to determine their configuration relative to one another and to determine the response of each physical camera unit.

Algorithms designed and implemented on this kind of delivery platform must be robust and scalable, to allow for reliable and widespread deployment. Our first application in the context of the distributed camera array project is the Local Positioning System (LPS). LPS is a new technique for 3D position tracking using active LED markers detected using video cameras and processed with special purpose hardware. The system is designed to be scalable, accurate, and to have low-latency, for use in tracking the 3D location of a person in a localized area, such as a room or a building.

The LPS system constists of a set of two different kinds of components: video processing units, and mobile IR LED tags. Each video processing unit consists of a CMOS video camera used to capture a video stream, an FPGA which implements low level video processing and a ARM microprocessor which implements the high level tracking algorithm. An ethernet network interface connects each unit to one another and system controllers. The mobile IR LED tags contain a flashing IR LED, which flashes in a pattern indicating its unique ID. These IR transmissions are captured by the video processing units to localize the unique position of each tag.

Our current hardware prototypes onto cover about 20'x20' of space, depending upon the tracking resolution desired. To investigate the research question of scalability to larger spaces and larger numbers of video capture and processing units under various conditions could be accomplished by 1. massive hardware deployment or 2. with software simulation. The second option is presented here as a software solution for exploring the performance of this approach using different algorithms under varying conditions. This simulation approach may be used not only for the LPS system, but to investigate general calibration, performance, and scalability issues in a variety of computer vision applications. The simulation allows to the user to specify a scene description comprising a scene geometry, camera placement, camera distortion models, tag descriptions, tag motion, and hooks for implementing a variety of processing architectures.

Context-Aware And Template Based Workflows

Gerald Tarcisius, M.A.Sc Student, Prof. Kostas Kontogiannis,
Electrical & Computer Engineering
University of Waterloo, Waterloo, Ontario, N2L 3G1

Over the past few years the Internet has evolved from a global repository of static, read-only web pages into a distributed computing platform. Specifically, it provides the ability for businesses to interact with each other as well as with their consumers via the use of remotely invocable applications known as Web Services [1]. Furthermore, complex Web Service invocation scenarios, tasks, and processes can be modeled and enacted by using customized workflows.

For this research we design, a meta-model and a novel framework that associates context-awareness to the run-time dynamic instantiation of workflows according to specific quality and user requirements. A workflow is a generic mechanism for describing a business process that is modelled as an Event Condition Action script. A segment is a mechanism for describing common context characteristics for a group of services. A service is a means for describing functionality delivered as part of a workflow and is modelled as a node in a workflow graph. Similarly, a stream is a specification for connecting services to each other in terms of their exchanged data. A template of an entity is an abstract parameterized class that can be instantiated at runtime into a concrete typed instance of that entity. An entity may either be a workflow, segment, service, or stream. In our work, Context refers to any information that can be used to characterize the situation of an entity [2]. Context-awareness refers to the use of context to provide task-relevant information or services to a user. Templates are instantiated as concrete entities based on some ontology requirements, and concrete entities are given concrete groundings based on the specific context requirements of the entity.

The concepts of personas, roles, and interaction spaces are also introduced in the meta-model to complete the framework. A persona is a collection of roles an agent can take. A role is a grouping mechanism that is used to categorize personas based on various properties or behaviours. An interaction space is a mechanism for controlling and coordinating the instantiation and the enactment of workflows. The templates are instantiated by applying a Web service localization and selection mechanism that allows for the identification of the service that optimally matches the context and the quality of service requirements set by the template specification. The matching process utilizes the comparison of feature vectors, ontologies, as well as, personas, and invocation profiles. The prototype system is being implemented in collaboration with IBM Canada, Center for Advanced Studies.

[1] Michael Champion, Chris Ferris, Eric Newcomer, and David Orchard. Web services architecture. World Wide Web, http://www.w3.org/TR/ws-arch/.

[2] G.D. Abowd and A.K. Dey. Towards a better understanding of context and context-awareness. In Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing, pages 304-307. Springer-Verlag, 1999.

Incremental Utility Elicitation with the Minimax Regret Decision Criterion

Tianhan Wang, Master's Student
Department of Computer Science
University of Toronto

Supervisor: Dr. Craig Boutilier

Utility elicitation is a critical function of any automated decision aid, allowing decisions to be tailored to the preferences of a specific user. However, the size and complexity of utility functions often precludes full elicitation, requiring that decisions be made without full utility information. Adopting the minimax regret criterion for decision making with incomplete utility information, we describe and empirically compare several new procedures for incremental elicitation of utility functions that attempt to reduce minimax regret with as few questions as possible. Specifically, using the (continuous) space of standard gamble queries, we show that myopically optimal queries can be computed effectively (in polynomial time) for several different improvement criteria. One such criterion, in particular, empirically outperforms the others we examine considerably, and has provable improvement guarantees.

Service Invocation And Roaming In Pervasive Computing Environments

Alvin Chin, Kostas Kontogiannis
Software Engineering Group,
Electrical and Computer Engineering,
University of Waterloo, Waterloo, ON, N2L 3G1

Computing has evolved from its initial mainframe centralized paradigm, to a highly decentralized and distributed mobile computing paradigm. Emerging mobile communication protocols aim to disassociate computing platforms from the location and the medium the user is accessing services. Moreover, a user should be able to access his/her personalized services any time, anywhere, on any type of device and with any type of wireless technology in a non-obtrusive environment. The complexities and problems associated with a mobile wireless environment should be insulated from the user. This current new era of computing is called pervasive computing [1]. With the vast number of different devices out there today like PDAs, cell phones, embedded systems and with all the different Operating Systems, messaging protocols, middleware standards, and presentation platforms, the challenge is how to easily enable them to access in a transparent and ubiquitous manner location aware Internet Web Services.

In this research such a framework for accessing location-aware services (service invocation) and the protocols involved for a user to be able to access services across different locations and networks (service roaming) are presented. Service invocation involves having to do service registration, service brokerage and service discovery, before service invocation. The proposed architecture uses W3C endorsed standards for web services. Specifically, WSDL and UDDI are used for service advertisement and registration, UDDI is used for service discovery and SOAP for service execution. There are many methods to achieve roaming at different layers of the network OSI model.

In our architecture, we propose tackling this problem at the application layer to allow for maximum flexibility and minimum system architectural modifications. Existing web-based applications rely on the HTTP protocol, which itself is a stateless protocol. In order to determine the particular services that a user is accessing and remember the task that the user is performing, we need to create a protocol that involves keeping track of state. This is done by creating a session associated with each user that logs into the system. To accomplish roaming, the architecture introduces the concept of a proxy which intercepts client requests and server responses, and stores them [2]. If the client crashes, is disconnected, or moves to a different location, the existing request/response can be easily retrieved or resumed. Currently, a prototype system for the proposed framework is being implemented in collaboration with Bell Canada, and the Bell University Laboratories at the University of Waterloo.

[1] Vince Stanford, “Pervasive Computing Goes To Work: Interfacing to the Enterprise”, IEEE Pervasive Computing, July-Sept 2002, pp.6-12.
[2] Bin Yao and W. Kent Fuchs, “Proxy-based Recovery for Applications on Wireless Hand-held Devices,” Proceedings of the 19th IEEE Symposium on Reliable Distributed Systems (SRDS'00), October 16-18, 2000, pp. 2-10.

Quality and Constraint Driven Workflow Composition

Faryaaz Kassam, M.A.Sc Candidate
Electrical and Computer Engineering
Software Engineering Lab (SWEN)
University of Waterloo, Ontario,

Complex business processes are becoming crucial for day-to-day activities and are vastly used in today's industry. Moreover, different companies involve different and customized models for conducting their businesses such as providing a service, manufacturing a product, purchasing, maintaining and updating critical information. As companies merge or embark in collaborations, there is a need to modify, reuse, and integrate existing business processes (workflows) and run-time applications. However, it is important that these newly composed workflows reuse as much as possible from existing ones. In order to achieve this objective, a mechanism for analyzing, updating, and integrating existing workflow specifications to generate new ones that conform to specific target requirements is needed.

In this research, we propose a matching process that yields the optimal composition of target workflows from constituent ones based on two different criteria. First, in order for a target workflow to be composed, we need a mechanism to compare services from the constituent workflows and to decide whether they match the requirements set for the corresponding services in the target workflow. For this we propose the use of Quality of Service (QoS) attributes for Web Services [1], ontologies, and service interfaces to determine whether two services match or can be composed. Specifically we propose a service equality framework where we consider that two services are equal if their QoS attributes are within a given range, are part of the same ontology, and their service interfaces are compatible based on their type, category, invocation protocols and so on. Second, during the composition process there might be certain compositional, structural, or behavioural constraints imposed on the target workflow. We propose the use of tolerance thresholds similar to those in constraint satisfaction problems [2] to restrict the number of constraints violated. Furthermore, we propose a workflow composition algorithm to produce the target workflow.

Such an algorithm to compute the optimal composition of workflows with respect to a target workflow uses a breadth first traversal to create the search space of alternative options, and simultaneously discards all irrelevant branches through the use of an optimal search algorithm. A variation and a refinement of the algorithm that involves incremental composition based on workflow segments, ontologies, and more detailed semantic descriptions of service distances can produce even more accurate result in terms of guarantees of the composition optimality.

[1] A. Main, A. Nagarajan, "Understanding Quality of Service for Web Services", IBM developerWorks, January 2002.

[2] R. Bartak, "Modelling Soft Constraints: A Survey", Neural Network World, Vol. 12, Number 5, pp. 421-431, 2002.

Needle Insertion Modeling for the Interactive Simulation of Percutaneous Procedures

S. P. DiMaio and S. E. Salcudean
Robotics and Control Laboratory
Electrical and Computer Engineering
University of British Columbia

Some of the most common procedures employed in modern clinical practice involve the subcutaneous insertion of needles and catheters. Such procedures range in complexity from superficial needle pricks to the biopsy of deep-seated tumours, and involve the subcutaneous insertion of long, slender surgical tools and needles into soft, inhomogeneous tissue, usually without visual feedback from below the skin's surface. Physicians and surgeons often rely only upon kinesthetic feedback from the tool, correlated with their own mental 3-D visualisation of anatomic structures. Difficulties arising from the complexity of such interventions have been studied in biopsy, brachytherapy and particularly in anaesthesia, where it is found that complications are due, in large part, to poor technique and needle placement.

Virtual-reality-based training systems are of great interest for simulated training and planning. Much of the existing work on virtual environments for needle insertion has not been validated and is not generalisable. Some of this work is based on needle driving forces measured in gelatine, ex vivo porcine and bovine tissues; however, forces are measured at the base of the needle only, while in fact penetration forces are distributed along the entire length of the needle axis, due to several phenomena including cutting, friction and tissue deformation. These approaches may not be suitable for training simulators involving complex soft tissue anatomy, needle placement optimisation, trajectory planning and automatic control, where more detailed verifiable knowledge of the insertion biomechanics is required. The objective of this work has been to determine the forces occurring along the needle shaft during insertion, and to use these in physically-based needle insertion simulators, trainers and planners.

A novel interactive virtual needle insertion simulation is presented. The insertion model simulates three-degree-of-freedom needle motion, physically-based needle forces, linear elastostatic tissue deformation and needle flexibility for the planning and training of percutaneous therapies and procedures. To validate the approach, an experimental system for measuring planar tissue deformation during needle insertions has been developed and is presented. A real-time simulation algorithm allows users to manipulate the virtual needle as it penetrates a tissue model, while experiencing steering torques and lateral needle forces through a planar haptic interface. Efficient numerical computation techniques permit fast simulation of relatively complex two-dimensional and three-dimensional environments at haptic control rates.

Simulating Surgical Cutting

Hui (Hilary) Zhang, Master Degree Student
Experimental Robotics and Computer Graphics Laboratory
Simon Fraser University, Burnaby, BC, V5A 1S6

Surgical simulation is a promising technology both for helping to train surgeons and for surgical planning. An important capability for any surgery simulation systems is the ability to generate realistic cuts through soft tissue models. We believe a key characteristic for virtual environments is sufficiently realistic, real-time visual and haptic rendering of the simulation of cutting tasks. This way, trainees can graphically visualize the outcome of a cutting task while feeling reaction forces which arises during their action, through a haptic user interface device. We found that a surface mesh model offers an adequate representation of the objects. We use an enhanced surface mass-spring model to simulate virtual cutting operations, using an input device with haptic feedback. We introduce novel algorithms to subdivide the surface and generate interior structures progressively, i.e. in a way that follows the motion of the input device with almost no delay. Two types of cutting, cut-into and cut-through, are supported in our simulator. These cutting techniques are coupled with a force feedback (haptic) device and graphic visualization will be integrated into a training environment for both open and minimally invasive surgery.

Video Analysis For An Endoscopic Bimanual Task Under The Influence Of Camera Rotation And Display Location

Bin Zheng, Farah Verjee, Christine L. MacKenzie
Human Motor Systems Laboratory,
Simon Fraser University, Burnaby, B.C. V5A 1S6

During minimally invasive surgical procedures, the surgeon may be required to operate with the endoscopic camera rotated along the longitudinal axis of the endoscope. Consequently, this camera rotation causes the displayed working field to be misaligned with the actual working field. The majority of tasks during these surgeries involve bimanual movements therefore, our study investigates the effects of camera rotation (0 degrees vs. 45 degrees) and display location (vertical vs. superimposed) when performing a reach, grasp, and cut task with both limbs. Past studies showed the effects of camera rotation are more pronounced with increases in task difficulty (Zheng et al., 2003). As the coordination of surgical instruments within each hand is a relatively complex task, we presume that the effects of camera rotation will degrade task performance. Furthermore, we expect that a superimposed display would alleviate these effects of camera rotation on task performance, compared to the vertical image display.

Eight subjects performed the same reach, grasp, and cut task in a mock surgical set-up that included an endoscopic camera and training box. The image of the work plane inside the training box was projected either on a vertical monitor at subjects' eye level or superimposed over the training box by means of a half silvered mirror. The work plane consisted of an embedded thread within a synthetic tissue (fake ear). The right-handed subjects were instructed to use a laparoscopic grasper held in the left hand to reach, grasp and pull out the embedded thread a short distance. Then they were required to cut the thread beneath the grasper with a pair of endoscopic scissors held in the right hand. Subjects initially performed three practice trials to become familiar with the movement. Five trials in each of the four conditions were performed, thus data from 20 trials was collected.

Each trial was videotaped and a number of important events from the movement of the tips of the grasper and scissors were identified. From these events, the entire task was subdivided into subtasks required to reach the final goal of cutting the thread. Dependent measures included durations between these identified events. Duration data were submitted to a 2 (image display: vertical, superimposed) X 2 (camera rotation: 0, 45 degrees) X 5 (trials)) repeated measures analysis of variance.

Significant differences were found for total time to complete the task, between the 45 degree camera rotation and the 0 degree camera rotation, indicating that task performance degraded when the displayed and working field were misaligned. As the trials progressed, there was a significant decrease in the time needed for the grasper to reach over to the thread (i.e., a decrease in grasper reaching time). Display by trial interactions were also revealed for total time and grasper reaching time. When the working plane was vertically displayed, there was no change over trials; however, when the image was superimposed, total time and grasper reaching time became shorter as the trials progressed.

Communication of proprioceptive information between the two effectors is discussed in the context of remote manipulation. We outline the role of proprioception during a bimanual task and the benefits to task performance when the camera is rotated. Implication are addressed for optimal endoscopic procedures and layout designs.

[1] Zheng B, Janmohamed Z, MacKenzie CL (2003) Choice Reaction Time in an Endoscopic Task. Surgical Endoscopy. (in press)

Acknowledgements: This research was supported by IRIS, NSERC and MSFHR.

EXTENSIONS TO METRIC-BASED MODEL SELECTION

Yoshua Bengio and Nicolas Chapados

Metric-based methods have recently been introduced for model selection and regularization, often yielding very significant improvements over the alternatives tried (including crossvalidation). All these methods require unlabeled data over which to compare functions and detect gross differences in behavior away from the training points. We introduce three new extensions of the metric model selection methods and apply them to feature selection. The first extension takes advantage of the particular case of time-series data in which the task involves prediction with a horizon *h*. The idea is to use at *t* the *h* unlabeled examples that precede *t* for model selection. The second extension takes advantage of the different error distributions of cross-validation and the metric methods: cross-validation tends to have a larger variance and is unbiased. A hybrid combining the two model selection methods is rarely beaten by any of the two methods. The third extension deals with the case when unlabeled data is not available at all, using an estimated input density. Experiments are described to study these extensions in the context of capacity control and feature subset selection.

Recognising Panoramas

Matthew Brown
Laboratory for Computational Intelligence
University of British Columbia

This problem considering in this work is the fully automatic construction of panoramas. Fundamentally, this problem requires recognition, as we need to know which parts of the panorama join up. Previous approaches have used human input or restrictions on the image sequence for the matching step. In this work we use object recognition techniques based on invariant local features to select matching images, and a probabilistic model for verification. Because of this our method is insensitive to the ordering, orientation, scale and illumination of the images. It is also insensitive to 'noise' images which are not part of the panorama at all, that is, it recognises panoramas. This suggests a useful application for photographers: the system takes as input the images on an entire flash card or film, recognises images that form part of a panorama, and stitches them with no user input whatsoever.

[1] R. Szeliski and H. Shum. Creating Full View Panoramic Image Mosaics and Environment Maps. Computer Graphics, Vol 31, p251-258. 1997.

[2] R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press. ISBN 0521623049. 2000.

[3] D. Lowe. Object Recognition from Local Scale-Invariant Features. In Proceedings of the International Conference on Computer Vision, p1150-1157. 1999.

[4] M. Brown and D. Lowe. Invariant Features from Interest Point Groups. In Proceedings of the 13th British Machine Vision Conference, p253-262. 2002.

Tele-Robotic Management System

Jamie King, Ray Pretty, Brendan Brothers. and Ray Gosine

This poster illustrates an application called the Tele-Robotic Management System (TMS) for coordinating multiple operators with multiple robots for applications such as underground mining. TMS utilizes several graphical interfaces to allow the user to define a partially ordered plan for multiple robots. This plan is then converted to a Petri net for execution and monitoring. TMS uses a distributed framework to allow robots and operators to easily integrate with the application. This framework allows robots and operators to join the network and advertise their capabilities dynamically through services. TMS then decides whether tasks should be dispatched to a robot or a remote operator based on the services offered by the robots and operators.

Language Modeling

Maryse Boisvert
Informatique et recherché opérationelle
Université de Montréal

A goal in statistical language modeling is to learn the probability distribution of sequences of words. In a recent approach proposed to model the word sequence probability, Bengio (2000) introduced a novel neural network architecture and obtained significant performance improvements over traditional N-gram models. The model presented here proposes to extend this algorithm to include semantic information (the word senses in the widely used WordNet database) in the learning algorithm, resulting in a disambiguation table and a language model based on senses.

The model expresses the joint probability of a word sequence in terms of the conditional probability of a word given its sense, along with the probability of a sense given the last N senses. The probability function of a sequence of senses is learned by the neural network along with distributed feature vectors for each sense. A Hidden Markov Model is used to represent the possible sequences of senses given a sequence of words, and the disambiguation table is estimated via the EM algorithm. The distributed representations obtained on the senses allow generalization, since a sequence of senses gets a high probability if it is similar (in terms of feature vectors) to a sequence of senses used to form an already seen sentence.

References:

Y. Bengio, R. Ducharme and P. Vincent, “A Neural Probabilistic Language Model”, Technical Report 1178, Département d’informatique et de recherche opérationnelle, Université de Montréal, 2000.

CaML: Camera Markup Language for Network Interaction

Jeffrey E. Boyd (Asst. Professor)
Nathan Kendrick (MSc Student)
Department of Computer Science, University of Calgary
Calgary, Alberta, Canada T2N 1N4

As processor speeds increase and the cost of digital video technology falls, the use of video is expanding in a plethora of applications. These applications include video surveillance for security applications, group awareness systems in which video systems facilitate group interactions over a network, tele-instruction where video and sound are the essential media by which a lecturer communicates with students over a distance, and enhanced broadcasts that use information extracted from a video stream to add information to, and thereby enhance, the broadcast content of a public events (most often sports).

A major problem that now faces developers of video systems is the requirement to build the low-level video processing from the ground up for each application. For example, these video systems usually do computations for segmentation, connected components, object features, recognition and, in some cases camera calibration. Furthermore, some applications require that a user of the system be able to control pan, tilt, and zoom, and image quality and format. Video system development would be easier and faster if these basic tasks could be removed from the system design process so that each developer would not have to implement them.

We describe a camera system that acts not merely as a provider of pixels, but as a video information server. The system integrates the functions of a camera and the low-level tasks required by a video system, and then serves the resulting data to client applications. Thus, a video application need not deal with low-level pixel data and algorithms, but can be built starting from the level of image object information, retrieving its input from the information server. An application interacts with the camera server using the Camera Markup Language (CaML, pronounced camel) proposed here. CaML is an XML-based (Extensible Markup Language) data format for exchanging video information with a server. Although CaML refers to the language of interaction, casually we also use the term to refer to the server.

CaML provides a layer of abstraction between the application and the pixels to simplify the process of developing video applications. It is well-suited to exchanging data over a network because of its XML roots. Using a camera as a server on a network makes it a simple matter for a single application to use multiple cameras. Local- and wide-area networks (LANs and WANs) replace the need for conventional methods for routing video signals.

We have demonstrated CaML servers in applications that monitor the activity of people working in a laboratory environment. Also, a CaML server provided the motion feedback in a new-media art display called Swarm Art (swarmart.com) in which swarms of artificially intelligent boids move and react to the motion of observers. Using CaML, the creators of Swarm Art quickly and easily added the video capability to their system to monitor the motion of observers in front of the display, without any expertise in video processing.

A Study On Using Biometric Sensors For Detecting User Emotions.

Cristina Conati, Heather Maclaren, Romain Chabbal
Department of Computer Science
University of British Columbia
Vancouver, BC, V6T 1Z4
conati, maclaren@cs.ubc.ca

Assessing the emotions of a user within a real-time environment is an important aspect of affective computing [2] that can greatly enhance the capability of an adaptive user interface to provide effective, unobtrusive feedback and determine the most valuable approach to use on each occasion when interaction is required.

How detailed the information collected should be depends on the task and the application, going from a general overview of emotional valence and level of arousal to a single specific emotion, to multiple specific emotions. Clearly, the more detailed the emotional information required, the more uncertain and prone to error the modeling task becomes. The long-term goal of our research is to devise a framework for affective modeling that can detect multiple specific emotions. To handle the high level of uncertainty in this modeling task, the framework is designed to integrate in a Dynamic Decision Network information on both the causes of a user's emotional reactions and their effects on the user's bodily expressions [1]. We have performed an empirical study
designed to understand how these effects can be monitored in real-time by using affective biometric signals [2].

To date, the assessment of emotions from biometric signals within a real-time environment has been attempted only in quite constrained contexts such as the detection of a specific emotion (e.g. level of anxiety). Biometric sensors have also been used to detect several different emotions deliberately expressed by a professional actress [3]. Using biometric sensors for real-time recognition of multiple emotions, expressed spontaneously in an uncontrolled environment, has not been previously attempted. Thus, it is not clear how effectively the sensors can detect emotions that may be expressed more subtly, that could interfere with each other and whose duration is unknown, in an environment with a possibly higher level of noise due to motion artifacts. To shed light on this issue, we have conducted a user study in which four biometric sensors were used to detect the emotions of students playing with Prime Climb, an electronic educational game. We give an overview of the study performed and illustrate the difficulties encountered due to operation in an uncontrolled environment. We present the observations made of the monitoring process itself, and initial findings whilst attempting to process the collected data in a real-time manner.

[1] Conati C. Probabilistic assessment of user's emotions in education games. Journal of Applied Artificial Intelligence, special issue on Merging Cognition and Affect in HCI, 2002.

[2] R. Picard. Affective computing, 1995.

[3] Picard R.W, Vyzas E., Healey J. 2001. Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. IEEE Transactions on Pattern Analysis and Machine Intelligence. 23(10).