Research Plans

My research lies at the intersection of works of G. Brownell, the late T. Cheatham, A-L. Brownell, D. Chesler, B. Horn, R. Brooks, J. Shah and D. Mumford and has lead to several broad, but related research agendas. My research program reflects the essential interplay between theory and practice of computer science and machine vision.

One principal component of my work has been on the design and implementation of novel algorithms and visualization techniques on parallel and distributed computers and on high performance multimedia systems. Real-life applications have included new imaging systems and hardwares dedicated to studying brain diseases, neuroscience, and autonomous navigation. I focus also on how these new quantitative methods are connected to user interface design (human/machine interaction) and expert systems for interactive multimedia. This is accomplished using new mathematical and artificial intelligence techniques from differential geometry, constrained optimization, integral equations, piecewise differential equations, image analysis, data mining and machine learning as well as accurate modelling of data acquisition to truly reflect the physical principles.

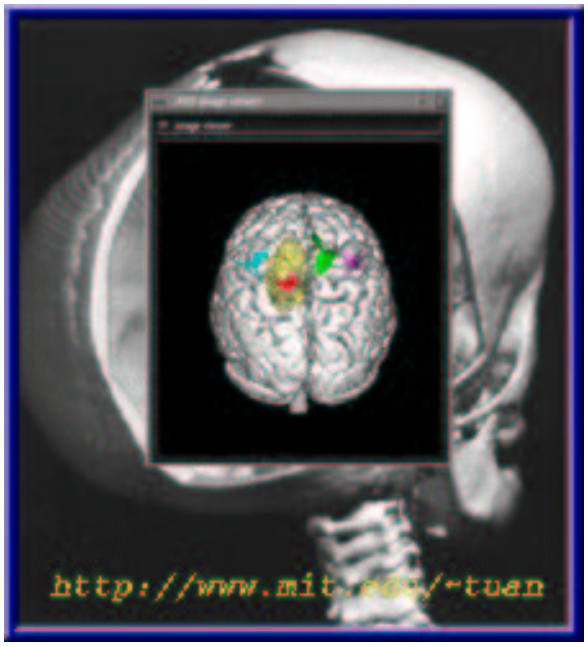

Figure 1: We think of the problem of 3D reconstruction as that of functional minimization. 3D visualization for minimal post-operative neurologic deficits: Yellow is tumour volume.

Interestingly, in several instances my labours in a specific problem turn out to have a clear "*dual use*" impact in other application areas. For example, results from our fast registration and recognition methods (data fusion and object recognition by a machine) that support visualization of 3D structures that were originally developed to provide more accurate diagnosis and maximal tumour resection, could also be relevant for other situations such as in the area of *multisensor sensor fusion* that is essential in autonomous navigation of aircraft, vehicles, and robots. This has so resulted in interests and contacts from researchers with the Mars Lander Space Programs on "terrain-aided aircraft navigation" (US/Canada) and a large Japanese car manufacturer wishing to develop a program in "Intelligent Transportation". I have applied for two US patents; one of which is licensed to a small US company.

I have tried to give credits to my co-workers in this "statement of research plans". But there are certainly omissions because of the time constraints. This will be corrected in the next version and a reference section will also be added as well.

Contents

1 Error Models for Recognition, Estimation and Tracking

The Problem: When do current object recognition, estimation and tracking systems break? Where are the key hard parts of the recognition, estimation and tracking problem? This project seeks a framework in which to answer efficiently these questions.
Motivation: Most current recognition and tracking systems have been tested on a small number of examples, with no careful study of how well they will work as the problem domain gets harder. For many scientific applications, one is interested in obtaining not only the estimates but also statistics for the errors in the estimates, which allow one to quantitatively evaluate the quality of the estimates in different locations. Although, there have been many approaches to the problems of recognition, estimation and tracking, few, if any, address the issue of error statistics. Furthermore,

many of these systems blithely assume that uncertainty in the image data can be ignored, or treated as a simple distribution. We argue that it is dangerous to make such assumptions about the effects of uncertainty on a recognition system, and that it is important to predict degradations in the system as a function of problem parameters.

Approach: We assume that image features, such as edges and vertices, are measurable only to within some bounded uncertainty, represented as an ε-disc. We then consider the effects of such uncertainty on the computed transformation relating an object model with an image. This typically leads to a set of feasible transformations. We have so far primarily focused on methods for recognizing objects in cluttered, noisy, unstructured environments. Such systems have been been incorporated as part of a hand-eye system, as part of a navigation system for autonomous vehicles, and as part of an inspection and process control system for industrial parts. We have focused also on formal methods for evaluating alternative recognition methods, on developing provably correct matching schemes for recognizing objects, on grouping methods for preprocessing the input data into salient sets of features, on the role of visual attention in recognition, and on efficient methods for indexing into large libraries of objects.

Our work concerns also with a novel filter-based, multiresolution approach to velocity estimation allowing coarse-to-fine refinement of the velocity estimates. It is based on an affine model to describe local motion variation within a sequence; and the incorporation of this local model into a multiresolution framework to describe the global motion field. The algorithms are efficient and lead directly to precise descriptions of motion vector field. Experiments have shown that the estimator and tracking mechanism performs well on both synthetic and natural sequences. Shown in Figure 2 is such a study of the dynamic mature human heart.

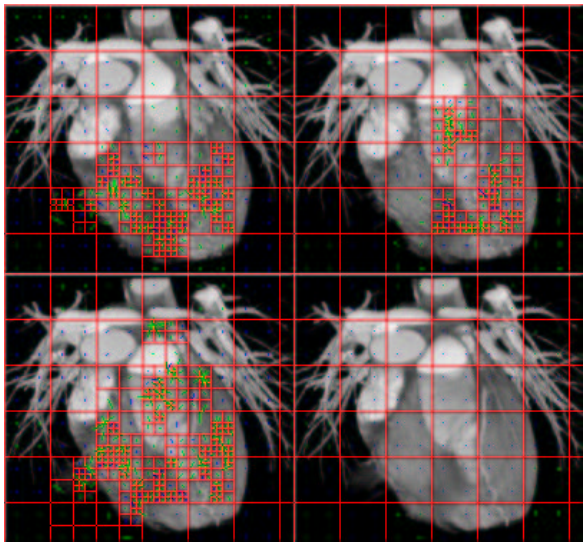

Figure 2: "Multiresolution estimation and tracking of the mature heart" In preparation for submission to NATURE.

Applications of our results have included active head-eye vision systems and the problems of image guided surgery and enhanced reality visualization, as shown in Figure 5. One novelty of our approach is a mathematical framework and algorithms for automatic, fast, on-line recognition. Motion estimation, tracking and object recognition are key components in a large number of applications such as in autonomous navigation, video data compression, surveillance, image understanding (motion-based segmentation, depth from motion), image registration and compositing. The first step in processing image sequences is typically image velocity estimation (optical flow field). Much research effort have been spent in these areas, but designing general and robust solutions to these problems has proved difficult due to the complicated relationship between the motion of objects in a 3-D scene and the apparent motion of brightness patterns in a sequence of 2-D projections. Information about the relative depths of objects is lost in the projection, and observed motion in the image plane can result from other phenomena than object motion in the scene, such as changes in the intensities. Moreover, measuring the motion in a 2-D image

sequence is in itself a non-trivial task, complicated by, for instance, the presence of observation noise, occlusions and temporal aliasing. We use a formal characterization of this set to determine the sensitivity of the recognition method. Specifically, we use a combinatorial occupancy model to determine the likelihood of false positive identifications, as the problem parameters change.

Impact: Using such methods, we have been able to analyze several existing recognition methods, including Hough Transforms, Geometric Hashing, Alignment and Interpretation Trees. The results allow us to compare directly the performance of such methods, to predict when such methods will break down (such predictions have been verified independently on real data), and to determine where the key parts of the recognition problem lie.

2 Image Understanding (IU) in Information Technology (IT) with Emphasis in Medical Domains (MD)

2.1 Objectives

The goal of this initiative is to merge current and emerging techniques from the image understanding community with existing opportunities in surgical, diagnostic and clinical problems in medical imagery. Advances in health care/medicine are clearly of importance to the general populace. Aspects of medical applications are of particular relevance to the military, especially techniques that enable rapid deployment of diagnosis, triage, and trauma care to the battlefield, without jeopardizing highly trained personnel.

Image Understanding (IU) is directly revelant to many such applications, including visualization, diagnosis, image compression (for effective transmission of data), and telepresence feedback. Applications of image understanding methods to medical problems can also have a clear "dual use" impact as well. Many of the same situations arise in cases of natural disasters, remote expert second opinions, and related problems. For example, registration methods that support visualization of 3D structures could be used for remote diagnosis and treatment.

2.1.1 Possible application areas

There has been an impressive surge in the capabilities of medical sensors in the past few years, especially in 3D sensors (e.g. MRI, CT, PET, SPECT). These sensors are providing practitioners with a wealth of accurate 3D data that often carries additional information about underlying material properties. Concurrently, noninvasive techniques (e.g. laparoscopy, endoscopy, arthroscopy) are also rapidly increasing in use and sophistication.

In parallel with biomedical applications, we are also actively pursuing visual recognition systems for the next generation human-computer interfacing, distributed data management architecture for intelligent multimedia information retrieval, and Bayesian networks for knowledge discovery in large image database systems. Both of these developments are examples of emerging opportunities for IU techniques, in large part because the combination of data sources and restricted operating environments directly requires methods for interpreting, manipulating and highlighting 3D data in a manner that effectively enhances the surgeon's field of view.

Making use of the calculus of variations, we has previously formulated the three dimensional reconstruction problem from cross-section as one of minimizing an energy functional with respect

to a pair of functions[1]. This approach resulted in 3D images with excellent details as shown in Figure 4.

Current medical sensors are capable of delivering very rich, and accurate information, in scenarios with minimal clutter. This fits nicely within the operating characteristics of current IU recognition and registration systems. At the same time, the objects that must be handled are often flexible, and of complex shape, so that we need to extend our methods to deal with new shape representations, and a broader class of transformations. Thus the medical domain may well provide the IU community with a rich source of new problems, which will help stretch our techniques to the next generation. At the same time, application of current techniques may well lead to near term successes, because of the advantages of constrained problem domains and rich data sources. Applying IU techniques to problems in the medical domain does more than just provide benefits for particular applications. It also provides a focused means of pushing out the performance and capabilities envelop of existing IU technology. Thus, there are exciting opportunities (that include the possibility of making an impact on a very broad scale) for using IU methods to leverage current capabilities in the following applications:

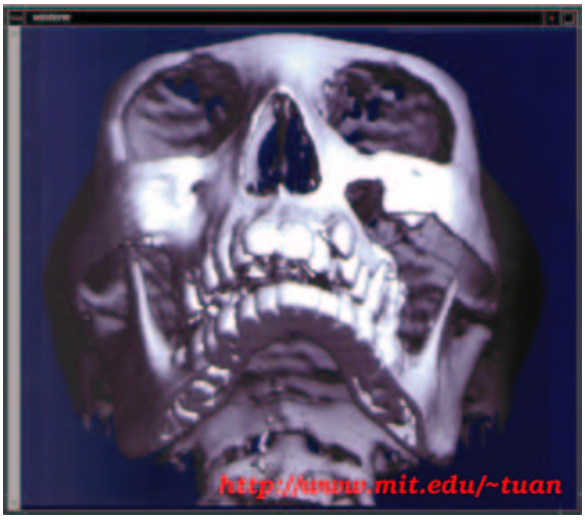

Figure 4: 3D visualization for planning surgery.

- 3D visualization of a patient's anatomy to support minimally invasive surgical procedures, remote diagnosis and telepresence surgery;
- realistic 3D simulation of anatomical structures and their mechanical properties, to support surgical simulations and planning of procedures, as well as the training of medical students;
- image guided surgical procedures, such as laser disc fusion, image guided biopsies, and focused ultrasound procedures;
- compression of medical data for transmission;
- enhanced clinical studies, by enabling new automatic registration and change detection of 3D medical imagery acquired over time, and by tracking anatomical structure through time sequences of imagery;
- medical database manipulation, especially retrieval and correlation of relevant data by image analysis and association;
- enhancing 3D medical data with traditional IU 3D data (stereo, shape from X, motion);
- 3D data fusion, especially multi-sensor fusion, such as among MRI, CT, SPECT, or PET, to support diagnosis and surgical planning;

[1]Joint work with D. Chesler (Harvard) and J.D. Chan, MD then at the Montreal General Hospital's Neuro-Radiology Department.

- quantitative descriptions of normal anatomy, especially the role of learning for building such descriptions from large numbers of examples.

2.2 Technical Issues

Because the range of medical applications is so broad, techniques from almost all areas of IU are relevant to problems in this area.

2.2.1 IU technologies needed

Many of the applications involve the registration of 3D medical data, such as MRI or CT, either to other medical data sets, or to 2d or 3D visual data of the actual position of the patient (enhanced reality visualization as shown in Figure 5). Such applications clearly benefit from robust and efficient registration and recognition methods, and the IU community has considerable experience in developing such methods.

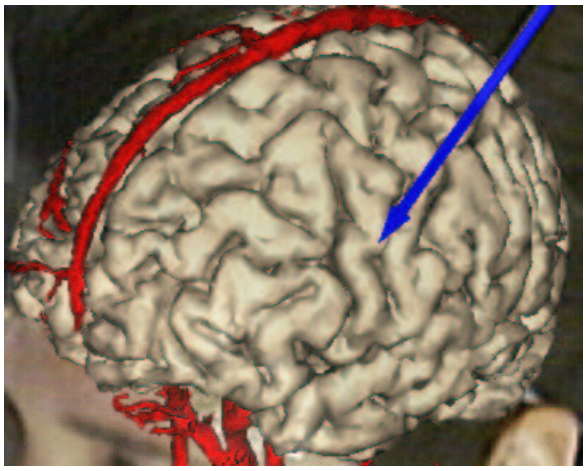

Figure 5: Enhanced reality visualization.

Similarly, clinical and surgical uses of medical imagery often require the segmentation of the data into distinctive tissue types. Again, IU methods for image segmentation are clearly relevant, especially those that use deformable models and those that can incorporate atlas information into the process. Similarly, clinical and surgical uses of medical imagery often require the segmentation of the data into distinctive tissue types. Again, IU methods for image segmentation are clearly relevant, especially those that use deformable models and those that can incorporate atlas information into the process. Other IU techniques that apply include:

- the representation of 3D shapes (especially deformable models) and generation and presentation of graphical displays of such shapes, especially for visualization tasks;
- motion tracking, in order to update the relative position of a 3D model with a surgeon's current viewpoint, and to update visualization capabilities, especially in minimally invasive techniques where the surgeon needs systems that can automatically expand the surgeon's limited field of view. See also Figure 2 and Section 1;
- change detection for clinical studies and to support diagnostic capabilities;
- shape extraction techniques like stereo vision and shape from shading, to augment medical 3D sensors;
- analysis of time series for perfusion studies, functional imaging, monitoring of surgical and drug treatments and motion analysis, such as cardiography or musculo-skeletal tracking;

- learning techniques, especially as applied to automatic segmentation, and to quantitative descriptions of normalcy;

2.2.2 Special concerns and constraints for IU in medicine

Perhaps the most critical factor for IU to have an impact in the medical arena is to ensure that the IU community establishes meaningful collaborations with medical practitioners. While IU clearly has potential contributions to make to medicine, they will have a realistic impact only with carefully coupled with actual medical practice. A number of techniques and systems have already been demonstrated in the medical domain by several US centers, including ours, and others in the world. A collaborative effort by MIT and MGH (Boston) have demonstrated a registration system that matches 3D data sets against one another. This system has been demonstrated as a frameless stereotaxy system for neurosurgery, by enabling an enhanced reality visualization of a video view of the patient from the perspective of the surgeon with an aligned view of an MRI or CT reconstruction of the patient's 3D anatomy. A similar frameless stereotaxy system, using X-ray views and MRI reconstructions has also been demonstrated by Stanford. Registration methods have been demonstrated for other applications. For example, our system has also been used to register MRI data sets taken at different times of the same patient, as part of a clinical study of MS. Such registration has been used to automatically highlight to the radiologist anatomical changes that have occurred between acquisition times. Such methods could be used for monitoring the effects surgical and drug treatment. Various groups are developing tools for surgical and therapeutic planning, based on the accurate, registered 3D models that can be built from medical imagery.

2.3 Benchmarking plans

The issue of appropriate benchmarking methods will depend on the specific applications. For example, for problems involving registration of data, benchmarking should include measures of accuracy of registration on standard data sets, range of conditions under which convergence to a correct solution occurs and with what probability, and measures of false local minima encountered. For problems involving segmentation, benchmarking can be done against data sets segmented by experts (many of which already exist), with different measures of missegmentation rates being most appropriate.

2.3.1 Where do we go from here?

First a personal remark. My work on "automatic object recognition" and "machine learning" has also been supported by the US Army. (The Principal Investigator who oversees the entire program is Prof. A-L. Brownwell.) It turns out that the results from our labours to solve one specific problem could also have an impact on other fields ("dual use" technologies). A small part of my work on "machine vision and learning" has been thus judged by this granting agency to have a potential military application (automatic recognition of targets or autonomous navigation of missile over an unknown terrain). So this part of my work remains privileged information. This is a result that I had neither expected nor sought. Given a choice, I will not pursue this research direction any further.

I believe that the convergence of computer science and "well tested" machine vision techniques with the developing medical sensor technology will lead to a very timely opportunity for merging

IU and medicine. Such a initiative, to be successful, must connect with the traditional medical research community, and with components of the health delivery system and Information Technology (IT). This can be achieved perhaps mostly through interactions with major medical research hospitals such as my collaboration with the Massachusetts General Hospital in Boston.

3 Novel approaches to singular value decomposition in Information Retrieval Technology

The diagram to the right indicates the software that we intend to develop in the boxes and the immediate applications to which the software may be applied in the ovals. The lines connect which software is applicable to which applications, though we recognize that our software is more widely applicable than these immediate applications. Our long term objective is to develop, analyze, and implement scalable numerical methods and software tools that may be used for the solution to nonlinear eigenvalue problems that arise in Information Retrieval Technology. We also wish to develop graph generation and computer graphics tools that may be used in the parallel solution of PDEs as well as some minimization problems.

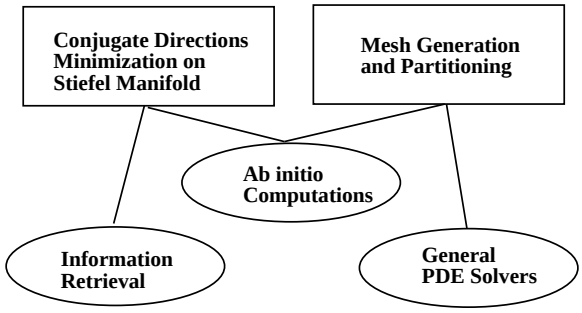

Figure 6: Our objective is to develop, analyze, and implement scalable numerical methods and software tools that can be used for the solution to nonlinear eigenvalue problems in Information Retrieval Technology.

3.1 Nonlinear Eigenvalue Problems as Minimization on the Stiefel Manifold

Berry describes four numerical methods for the computing the singular value decomposition of large sparse matrices associated with a method of latent semantic indexing described by Dumais, Furnas, and Landauer and Deerwester *et al*. The idea of their method is to search for a document based on being near a set of concepts, rather than using a traditional search based on keywords. A sparse term-document matrix A may be introduced by tabulating the number of times an important term appears in each document.

The singular value decomposition represents A as $A = \sum_{i=1}^{n} u_i \sigma_i v_i^T$, where $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_n \geq 0$. We may approximate this matrix by $A_k = \sum_{i=1}^{k} u_i \sigma_i v_i^T$. By taking k large enough to contain the essential information in A, but small enough to obtain a reasonable compression, we have a method for quickly obtaining useful responses to a search query.

Berry considers computing the singular values by solving either the eigenvalue problem for $B = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix}$ or the eigenvalue problem for $B = A^T A$ (or $B = AA^T$). One of his four methods is a trace minimization approach introduced by Sameh and Wisniewski.

Recent experience suggests that it is advantageous to reformulate the problem as a constrained optimization problem

$$\min X^T A(y(X))X, \tag{1}$$

where

$X \in R^{n,m}$	is the eigenvector matrix with orthogonal columns (the wave functions) i.e. $X \in$ Stiefel Manifold.
$y(X) \in R^n$	is a vector valued function of X (charge density)
$A(y) \in R^{n,n}$	is a symmetric matrix valued function of the vector y (the Hamiltonian).

Note that the matrix A depends on the eigenvectors.

The problem may also be roughly viewed as the nonlinear eigenvalue problem $A(y(X))X = X\Lambda$, where Λ is the diagonal eigenvalue matrix. Nonlinear eigenvalue problems have not yet been systematically studied by numerical analysts, though Broyden's method was explored for density functional self-consistent field problems. A number of authors have recently attempted conjugate directions style approaches or alternatives to the constrained minimization problem. We wish to improve on these methods by formulating the problem in the mathematically most natural manner – as a minimization problem on the Stiefel manifold. We then wish to develop software that will solve this problem on this manifold.

The Stiefel manifold $V_{n,k}$ is well known in mathematics to be the manifold in matrix space consisting of n by k matrices with orthonormal columns. When $k = 1$, this is the sphere, while when $k = n$, this is the entire space of orthogonal matrices. Minimization problems for linear eigenvalue problems may be efficiently solved directly on the Stiefel manifold by following geodesics right on the surface. On general manifolds, such geodesic following requires the solution of complicated systems of differential equations, but on the Stiefel manifold, following the geodesic may be performed with routine matrix computations. We wish to extend this idea to non-linear problems thus implementing a "proper" conjugate directions minimization approach.

3.2 Unstructured meshes

The simplest mesh for such a problem is the uniform regular grid in either two or three dimensions. For example, current production codes in use uniform grid based FFT method. Uniform grids are also standard in finite difference computations.

However, uniform grids have serious disadvantages: (1) an unnecessarily large number of grid points are used and (2) boundaries are not resolved smoothly. The first problem is well known to strain computational power and memory. The second problem introduces large discretization error in numerical approximation.

To have a better approximation of irregular physical problems, as the one we are solving, with few mesh points, it is important to use adaptive methods and hence unstructured meshes which allows us to solve much larger problem with available parallel machines.

Unstructured meshes, however, lead to new challenges: programming difficulty increases, efficiency of algorithms decreases over regular meshes of the same size, and the overhead on parallel machines can be significant. New parallel algorithms for both mesh generation and partitioning must be designed.

3.3 Geometric methods for graph partitioning and mesh generation

Problems arising from physical simulation usually have an underlying geometric structure. Thus their graphs are usually *meshes*, that is, graphs whose vertices are embedded at specified locations in space. Vertices near to each other in the graph are also near to each other in space, in some sense.

Mesh generation produces such a discretized approximation. Solving a large problem on a parallel computer with distributed memory usually requires that the data for the problem be partitioned somehow among the processors. The quality of the partition affects the speed of solution; a good partition divides the work up evenly and requires as little communication as possible.

3.4 Parallel adaptive mesh generation

Mesh generators accept a geometry based definition of the problem and produce valid and numerically stable meshes. The geometry based definition may come from the error analysis of the previous step in an iterative method or previous step of time series simulation.

The basic elements of a triangular mesh are triangles in two dimensions and tetrahedra in three dimensions. Not all triangulations, however, serve equally well; numerical and discretization error depend on the *quality* of the triangulations, meaning the shapes and sizes of triangles. A typical quality guarantee gives a lower bound on the minimum angle in the triangulations.

We propose to use quad-tree and oct-tree based method to generate well conditioned unstructured triangular mesh. Sequentially, it produces a well-conditioned mesh by adaptively refine proper boxes that forms a nested quad-tree (oct-tree) of boxes. In other words, a *quadtree* is a recursive partition of a region of the plane into axis-aligned squares. One square, the *root*, covers the entire region. A square can be divided into four *child* squares, by splitting it with horizontal and vertical line segments through its center. The collection of squares then forms a tree, with smaller squares at lower levels of the tree. It has been shown that the above approach generates a well-conditioned mesh whose size is very close to optimal. We propose to design efficient parallel mesh generators that can be used in our application and also in other application domains.

The theoretical mesh generator replaces sequential quad-tree refinement by parallel sorting. Using the information from the sorted sequence, the "basic frame" of the quad-tree can be produced in parallel. The final quad-tree is obtained by a parallel refinement of the basic frame, which requires a shallow depth point location structure on the basic frame. On shared-memory machines, such point location structured can be constructed efficiently and the point location has complex equal to the height of the data-structure. However, on a distributed-memory parallel machine, such methods for point location could be relatively expensive. We propose to use geometric sampling and the partitioning method below to speed-up this step.

3.5 Parallel graph partitioning

As shown above, meshes used in computation have rich geometric structures. A partitioner can use this geometric information to advantage in several ways. First, partitioning the graph can be reduced to partitioning a region of space (again, in a suitable sense). Second, using *geometric sampling*, the partitioner can work with a small randomly selected subset of mesh points but still generate a good partition with high probability. This makes the partitioner more efficient. Finally, most computational meshes are composed of elements (triangles or tetrahedra, for example) that are *well shaped*. Meshes of well-shaped elements are guaranteed to have good separators; this can be used to ensure the effectiveness of a partitioner.

Based on the above observation, working with X. Gu (Harvard), we have developed a geometric approach to mesh partitioning. Theoretically, our algorithm runs in linear time sequentially and $O(n/p)$ parallel time on a machine with p processors. It always finds a partition with provably good quality.

We propose to develop efficient, portable data parallel software for unstructured mesh partitioning based on our geometric approach and we have built a prototype in Matlab. We have performed various experiments on meshes used in practical finite-element methods. Our Matlab programs are already in data parallel format and hence it is easier to adapt the Matlab code to standard high performance parallel languages. We already have a parallel implementation. We will make further improvement and address the issues of scalability and portability in this project.

3.6 Software for conjugate direction minimization on high performance machines

An idealized conjugate direction method for the computation of the eigenvalues of a symmetric matrix would minimize trace(X^TAX) by staying within the Stiefel Manifold $XX^T = I_k$. Such a method may be generalized to solve any constrained optimization problem on the Stiefel manifold.

Our solution of the problem by conjugate directions, requires reformulating the idea of conjugate directions on a curved manifold. The basic concepts originate in differential geometry, but a practical understanding of numerical computing is needed in order for the computations to be efficient. We present a summary of the key ideas. the tangent space to the Stiefel manifold at a particular point may be represented as an n by k matrix whose top k by k square is anti-symmetric. For this representation to have meaning, we must choose a coset representative from the orthogonal group whose first k columns correspond to the point in the Stiefel manifold. Householder reflections are a convenient way to represent such a matrix with a minimum of storage.

Just as in flat space, new search directions are obtained by combining the gradient with the previous search direction in a manner that insures conjugacy with respect to the Hessian of the objective function. However, unlike in flat space, the previous search direction must be parallel transported from the previous point on the Stiefel manifold. Neglecting to carry out this operation can destroy the superlinear convergence by not taking into account the curvature terms that are inherent to the manifold.

We plan is to implement this algorithm on scalable distributed machines and high performance multimedia systems. so that they may be used in both the solution the solution of information retrieval problems and other applications of Information Technology (IT) and Image Understanding (IU).

4 Real-time 3D reconstruction on high performance network and multimedia systems

Real-time 3D model generation is an exciting new area of vision research made possible by new imaging sytems, high performance network and multimedia systems. In a near future, multicameras will become ubiquitous and thus streaming media will be a truly interactive experience. (Imagine real-time 3d models that offers viewers options to “see” action from the any “dynamic” perspective. Such a complete 3D and real-time system presents many challenges. at MGH, I have implemented a real-time 3D reconstruction prototype. My work has focused on calibration, silhouette extraction and model computation. Volume intersection is at the algorithmic core of my model.

4.1 Hardware

We use 8 computers each equipped with 2 frame grabber boards a piece. The computers are connected over the intranet via 100 megabit ethernet. We do not add any special lighting or a scene backdrop to make processing easier. Sixteen cameras are mounted around a volume approximately 3 meters wide, 3 meters deep and 2 meters high. The cameras are not constrained to be in any particular location, but are position so as to sample the space as uniformly as possible. Four cameras are mounted on the top edges of the structure, four are put in the middle of the corner posts and eight cameras are position on tripods evenly distributed on each of the sides. The cameras all approximately point to a central area in the middle of the volume the size of person. In general, more geometry can be recovered if the cameras on opposite sides do not mirror each other.

4.2 Software

We have implemented algorithms to to extract the image of the object from its background and to build up a low resolution voxelated visual hull geometric representation. We created a silhouette map data structure from which we can interpolate the extracted silhouettes. We performed silhouette clipping of the low resolution geometry to the high resolution interpolated silhouettes. We used the captured images as projective textures to apply to the clipped geometry. We use MPI (Message Passaging Interface) as the communication interface for transferring data over the network. Images are captured at a resolution of 320x240. Images are taken simultaneously by the sixteen cameras and sent to a central computer which computes the 3D model from the data using silhouette intersection. We also tested a new technique which computed a 3D volume without first computing the silhouettes. The bandwidth needed to transmit the complete uncompressed images to a central computer is not available on our intranet. In addition, the bandwidth needs increase linearly with the number of cameras. Typically, we can only achieve a maximum 2 megabytes per second throughput over the network. A color image from one camera consists of 230400 bytes. For sixteen cameras, one time instant consists of 3686400 bytes and at 30 frames per second we would need 110,592,000 byte throughput over the network which is roughly fifty times the maximum available. We discuss a smart method of compression which approximately sends under 2000 bytes per image. At 30 frames per second with sixteen images this is 960,000 bytes per second well under the maximum throughput level. An additional advantage to this scheme is that it also does not need to decompress the data at the other end saving additional computational resources. We use a 320x240 image whose resolution has been reduced from 640x480. We have reduced the resolution for two reasons. First smaller images result in faster computation time, but more importantly we want to get rid of the interlacing effects from the camera. Each frame that is picked up by the frame grabber, contains data from two 640x480 frames 1/60sec a apart. The data in each row alternates between coming from each of the two frames. Merging these frames, gives a 640x480 frame approximately every 1/30sec. If the subject moves within this time, the frame appears to have stripping in the areas of motion. So we decimated the image to 320x240 to overcome this problem.

5 Silhouette Intersection

Our 3D model computation is based on silhouette intersection approach. This approach requires that the foreground object be segmented out as the silhouette. The volume computation is then

determined by the silhouette boundaries from each of the contributing images. This approach has a basic limitation, however, because the volume computed from silhouette data will never exactly correspond to the "true" volume. With an infinite number of cameras, the volume will converge to what t has been termed the "visual hull". We employed two schemes for computing the volume. In the first scheme, each voxel is projected back onto each of the silhouettes. If the voxels project back onto regions in each image that only contain the silhouette, this voxel is marked as part of the volume. This scheme depends on the resolution of the voxelated space. The second method is comprised of intersecting cone volumes. Each cone is determined by its base, the silhouette on the image plane, and its center, the origin of that camera. Although this method avoids voxel resolution problem, in practice it is difficult and computational expensive to compute the intersection of the volumes analytically and it depends on the volume representation. The pixels on the silhouette boundary are discrete, so an interpolation scheme might be necessary to define the cone.

The construction of accurate silhouettes is crucial component for an integrated real-time 3d geometry system. Accurate models are essential for volume construction and calibration. Speed is essential for a real time system. It is well known that silhouettes are very important features in conveying shape information. When rendering synthetic objects, textures can used to give the impression of high resolution detail, even when a low resolution polygon approximation is used. But using current methods, a very large polygon count is still needed in order to give the impression of a high resolution silhouette.

I propose a new method for silhouette clipping where low resolution geometry is clipped against a high resolution silhouette description. Even though silhouettes are often the easiest features to detect in photographs, they are also underutilized in present image based rendering techniques. Even methods such as the visual hull construction typically use a voxelized spatial representation and are not able to utilize all of the silhouette resolution apparent in the input images. In our research, by maintaining an explicit representation of the silhouette of an object as seen in many views, we can achieve renderings with high quality.

5.1 Calibration

To obtain images, we used a Kodak DVC 323 camera which attaches to the UDP port of the PC. It returns low quality images, but at an acceptable frame rate. We have also been using a Kodak DC 260 digital camera. This camera takes high quality images, but takes about 20 seconds to refresh after each still image. We used a passive calibration arm from FARO technologies Inc. The arm is 6 feet in length, and mounted on a dedicated table. A camera is attached to the arm, and moved around under human control. The arm reports six degrees of freedom in a stream to an attached PC. The goal of calibration is to geometrically coordinate images taken from many cameras. For simplicity we model each camera as a pinhole camera. This camera model allows us more then enough accuracy given the quality and resolution of our cameras. For example, our silhouette computation potentially has errors of 1-2 pixels. This error offsets improvements that could be made with a more complex cameras model (and saves us computation by not needing to use the camera model that has additional parameters). The camera model can be described by seven parameters, three for location, three for rotation and one for focal length. The seven parameters of each camera must be found in respect to a common coordinate system. This problem is easy to describe but difficult to solve – due to its non-linear nature.

6 Reliable and fast autonomous navigation with techniques from Machine Learning, Estimation Theory, and Computer Vision

6.1 Overview

I have designed and implemented a system for real-time automatic recognition and autonomous navigation. My system combines several independent results in the fields of machine learning, autonomous agent navigation, and computer vision into a coherent framework to support reliable and efficient real-time navigation.

The three key components of our system are: (1) Fast algorithms for exploration of an unknown environment modelled as a graph; (2) Localization of an autonomous navigating in a known environment using landmarks; (3) Visual detection and recognition of landmarks. I have designed and implemented several new algorithms for exploring unknown environments in a piecemeal manner. In this work, the environment is described by a simple discrete model, an undirected graph, and the challenge is to find algorithms that search this graph correctly and efficiently. I introduced a new method of localizing a autonomous agent in an unknown environment by the aid of landmarks. In this research, the environment of the autonomous agent is a continuous 2D plane with landmarks at certain points in the plane. The autonomous agent 's sensors are not perfect; instead, sensor measurements generally include noise. I developed a new method to estimate the autonomous agent 's position reliably and efficiently in the presence of measurement errors. The third part, on "visual detection and recognition of landmarks,", I implemented an algorithm to provide fast, on-line recognition of landmark[2] in noisy scenes. Images of landmarks taken by a navigating autonomous agent are matched with transformations of image models stored in a database. In the field of object recognition, the idea of matching a model of an object with an image by exhaustive search is standard. However, a large number of parameters is necessary to mold the image for a parameter match. This makes the search space extremely large. So my approach is novel and can provide real-time recognition.

I address the problem of *detecting* a particular landmark in an image and the problem of *recognizing* which landmark is in an image. My algorithm does not only determine if a particular landmark is in the image, but also where in the image. The algorithm finds the shape and orientation of the landmark. This can then be used to estimate the position of the navigating autonomous agent. A landmark in an image I (see Figure 7) is defined to be recognized if it highly correlates with a replica image R of the landmark. This replica image R contains a replica of the landmark at the location of the landmark in the image I and zero brightness values elsewhere. This replica is a transformation of an image of the landmark that serves as a model and is stored in a database. Recognizing a landmark means finding the parameters that describes such a replica transformation which leads to the highest correlation between replica image R and input image I (see Figure 7).

The replica in image R is a transformation of the image of the sign stored in the database. Notice that the image of the sign stored in the database was taken at a different time and position than image I. Therefore, it is impossible to obtain a perfect correlation (which would result in a completely black sign in the difference image). The space of possible solutions of the recognition problem is extremely large, even if a priori a particular landmark is known to be in the image. The dimension of the search space is determined by the number of possibilities for position, shape, and orientation of the landmark. The number of possibilities for the position of the centroid of the

[2] US patent pending.

landmark in the image is $O(n^2)$ for a $n \times n$ image.

If the landmark is known to have a rectangular shape, the number of possibilities for the size of the landmark is also $O(n^2)$. It is assumed that there are typical orientations of the landmarks that represent them well. For example, a traffic sign is usually seen from its front. Therefore, it is assumed that there are only ambiguities in the rotation of the landmark in the image plane. Even with this assumption, the number of possible angles is still very large; since the image is discrete, we assume that the number of possible angles is $O(n)$. Thus, the size of the search space is $O(n^5)$ for an $n \times n$ image. For a typical image of size $n = 256$, the search space has a size of order 10^{14}. Therefore, an exhaustive search would take too long to find a good match between the replica of the landmark and the image. Instead, I adopt a randomized search which uses simulated annealing. The quality of the solution is determined by the value of an energy or cost function that correlates brightness values in image and in replica. I experimented with various cost functions and search parameters. My method has the advantage of combining brightness and spatial information and it can be successfully applied to the general detection of landmarks, even in presence of noise and other objects with same brightness content. In particular, my method handles environmental noise such as graffiti on traffic signs and measurement noise due to the camera system. I view the problem of recognizing an object in a image as the problem of estimating the parameters that describe the object in the image. By addressing the object recognition problem in the framework of estimation and detection theory, I investigate if I could find a use of the Cramer-Rao bound. The Cramer-Rao bound for object recognition parameters *may* be useful in computer vision. The Cramer-Rao bound gives the optimal resolution obtainable for a given parameter, *i.e.*, the minimum estimation error for this parameter. This is independent of the estimation method, and therefore, the Cramer-Rao bound is a simple, general and yet powerful tool. For example, if a particular estimation technique obtains the same variance as the Cramer-Rao bound, then we know this is the best variance that can be obtained with any method. However, if the Cramer-Rao bound is very large, no technique can obtain a good estimate of the parameter. Using the Cramer-Rao bound then avoids wasting time trying to estimate poorly constrained parameters, and also wasting time finding that these parameters are poorly constrained by trial and error with specific object recognition algorithms.

Figure 7: Automatic recognition for reliable navigation.

For my research on "Error Models for Recognition, Estimation and Tracking" (*i.e.* when my recognition algorithms will fail!), please refer to Section 1.