Image Understanding and Robotics Research at Columbia University

John R. Kender[1]
Peter K. Allen
Terrance E. Boult

Department of Computer Science
Columbia University, New York, NY 10027

CUCS - 493 - 89

Image Understanding and Robotics Research at Columbia University

John R. Kender[1]
Peter K. Allen
Terrance E. Boult

Department of Computer Science
Columbia University, New York, NY 10027

0 Introduction

Over the past year, the research investigations of the Vision/Robotics Laboratory at Columbia University have reflected the interests of its four faculty members, two staff programmers, and 16 Ph.D. students. Several of the projects involve other faculty members in the department or the university, or researchers at AT&T, IBM, or Philips. We list below a summary of our interests and results, together with the principal researchers associated with them. Since it is difficult to separate those aspects of robotic research that are purely visual from those that are vision-like (for example, tactile sensing) or vision-related (for example, integrated vision-robotic systems), we have listed all robotic research that is not purely manipulative.

The majority of our current investigations are deepenings of work reported last year; this was the second year of both our basic Image Understanding contract and our Strategic Computing contract. Therefore, the form of this year's report closely resembles last year's. Although there are a few new initiatives, mainly we report the new results we have obtained in the same five basic research areas. Much of this work is summarized on a video tape that is available on request.

We also note two service contributions this past year. The *Special Issue on Computer Vision* of the *Proceedings of the IEEE*, August, 1988, was co-edited by one of us (John Kender [27]). And, the upcoming IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1989, is co-program chaired by one of us (John Kender [23]).

0.1 Low-Level Vision

0.1.1 Polarization and Specularities

1. New methods for using polarization to segment specular highlights and to separate reflectance components (Larry Wolff [44, 49]).
2. New methods for classifying material surfaces into conductors and dielectrics using the polarization of specular highlights (Larry Wolff, and Terry Boult [45, 50, 51, 52]).

0.1.2 Image Warping

1. A survey of image-warping techniques (George Wolberg [40]).
2. A novel data structure and algorithm for warping to and from arbitrary shapes (George Wolberg [41, 39]).
3. A new, highly efficient, general method for achieving 2-D image warps by separating the transform into two successive 1-D warps (George Wolberg [42, 43]).

0.1.3 Optic Flow, and Rotational Motion

1. New, provably optimal algorithms for determining optic flow based on smoothing splines (Anargyros Papageorgiou, David Lee of AT&T Bell Laboratories, Greg Wasilkowski of the University of Kentucky [30]).

[1]This work was supported in part by the Defense Advanced Research Projects Agency under contracts N00039-84-C-0165 and DACA76-86-C-0024.

0.3.2 Representations of Space

1. A survey of algorithms for the representation of space and free-space path planning (Monnett Hanvey [19]).
2. An analysis of the complexity of efficiently updating of digital distance maps in dynamic, greater than 2-D, environments (Terry Boult [11]).

0.3.3 Theory and Practice of Navigation

1. An analysis of the complexity of topological navigation by landmarks, with applications to the design of sensors and robot instruction languages (John Kender, Avraham Leff, and Il-Pyung Park [24, 25]).
2. Systems issues in real-time robotic navigation (Monnett Hanvey, and Russ Andersson of AT&T Bell Laboratories).

0.4 Parallel Algorithms

0.4.1 SIMD Algorithms

1. Analysis of two novel and several existing algorithms for depth interpolation using optimal numerical analysis techniques (Dong Choi, and John Kender [14, 15, 16]).
2. A method for shape-from-texture based on distortions in image autocorrelation (Lisa Brown, and Haim Schvaytzer of Cornell University [12]).
3. Programming environments and image pyramid emulation for the Connection Machine (Hussein Ibrahim, Lisa Brown, and John Kender [22]).

0.4.2 Pipeline Algorithms

1. Grey-level corner detection in real time (Ajit Singh, Mike Shneier of Philips Laboratories [33, 34]).
2. An integrated system for real-time visual object tracking (Peter Allen [1, 4]).
3. New algorithms for motion perception in real time (Ajit Singh [35, 36, 37]).

0.5 Robotics and Tactile Sensing

0.5.1 Integrated Environments

1. Integrated environments (Peter Allen, Paul Michelman, Ken Roberts, Amy Morishima, and Steve Feiner [2, 5, 6]).

0.5.2 Multi-fingered Object Recognition

1. Haptic recognition via active exploration with a robotic hand (Peter Allen, Ken Roberts [3, 7]).

We now detail these efforts, many of which are documented by full papers in these proceedings. We also include short discussions of work in progress.

1 Low-Level Vision

We have explored three areas of low-level vision, and the results that we have obtained in each of them came via the careful exploitation of new equations, representations, or settings for standard, traditional problems.

1.1 Polarization and Specularities

Prior to this research, the segmentation of specular highlight regions, and the separation of reflected light into diffuse and specular reflection components, could only be solved on dielectric materials (insulators). Additionally, previous algorithms were sensitive to color. However, by using polarization information, specular

2. New algorithms for the smooth interpolation of rotational motions (Ken Roberts, and Kicha Ganapathy and Garry Bishop of AT&T Bell Laboratories [31]).

0.2 Middle-Level Vision

0.2.1 Physical Stereo

1. A unified theory of generalized physical stereo vision for the determination of several first and second order local surface properties (Larry Wolff [46, 53]).
2. A new method for determining local surface orientation from a continuous variation of photometric stereo, called "photometric flow fields" (Larry Wolff [47, 54, 55]).
3. A new invariant within two-camera stereo that allows the determination of the orientation of lines and surfaces in a manner insensitive to baseline measurement error (Larry Wolff [48, 56, 57]).

0.2.2 Regularized Surface Reconstruction

1. A critical study of regularization methodology (Terry Boult [8]).
2. Investigations into the stability and error properties of a new integrated stereo matching, surface reconstruction, and surface segmentation (Terry Boult, Liang-Hua Chen, and Mark Lerner [9, 13]).

0.2.3 Sensory Fusion

1. A new method for classifying textures based on the relative contributions of independent texture methods to a fused texture percept (Mark Moerdler [28]).
2. A method for fusing texture and stereo (Mark Moerdler, and Terry Boult [29]).
3. A working system, now in production, for the spline-based recovery of smooth oceanographic positional information from noisy, conflicting input (Terry Boult, and Barry Allen of Columbia University's Lamont-Doherty Geological Observatory [10]).
4. An initial reexamination of depth-from-focus, for possible use in fusing with stereo and/or texture (Terry Boult).

0.2.4 Shape from Dynamic Shadowing

1. A patent for shape from darkness, a discrete method for deriving surfaces from dynamic shadows (John Kender, and Earl Smith [26]).
2. A parallelizable, optimal algorithm for shape from continuous shadows (Michalis Hatzitheodorou, John Kender [20, 21]).

0.3 Spatial Relations

0.3.1 Representations of Objects

1. An elegant representation for lines in three-space (Ken Roberts [32]).
2. New, robust measures for the error of fit of superquadric models to range data (Ari Gross, and Terry Boult [17, 18]).
3. An investigation into efficiently-computable invariants that quickly relate reflectance information to certain classes of generalized cylinders (Ari Gross).
4. The design and initial implementation of a system to numerically recovery the parametric representations of volumes from multiple types of data, and multiple sensor types (Terry Boult).
5. New algorithms for efficient viewpoint planning (Dino Tarabanis, and Roger Tsai of IBM Watson Laboratory [38]).

regions can be identified on both metals and dielectrics on a per-pixel basis, without the use of a segmentation procedure, as long as a controllable polarizing filter can be placed between the camera and the object. Most ordinary light sources are unpolarized, but light reflected off dielectrics tends to be much more polarized and is more easily separated. A few minimal restrictions on the phase angle between light, object, and camera, must apply; these and other qualitative statements have be made quantitatively precise. Initial experimental evidence is very encouraging, on a variety of metals, insulators, and metallic and non-metallic paints and glazes [44, 49].

Closely related to this method--indeed, derived from the same equations--are new methods for classifying material surfaces into conductors or dielectrics by using the polarization of specular highlights. As with the segmentation algorithms, they depend on the empirical determination of the polarization Fresnel ratio. Originally developed for use with point sources, the methods have also been extended to allow their computation to be based more typical extended sources, such as fluorescent tubes [45, 50, 51, 52].

Because both classes of algorithms have the same initial front-end requirements, both can be run in parallel on the same image for simultaneous material classification and separation of reflection components. In addition, a third class of algorithms can exploit further relationships implicit in equations for the polarization of reflected light, in order to determine local surface orientation properties (discussed below, under "middle-level" vision). Thus, the theory unites three very different roles of early vision: object composition, object position and orientation, and environmental lighting and reflections. Under construction is an integrated set of vision algorithms that does these (and other) tasks, called POLARIS, for POLarization And Radiometric Integrated System.

1.2 Image Warping

Many imaging situations call for small local geometric corrections on the retina; many graphic situations call for large ones. Remote sensing, medical imaging, and television commercials with special effects all share the need to elastically deform images to some ground truth or some aesthetic demand. Done in software, image warping can be thought of as a reconfigurable lens system. The existing technology is extensive, but relatively slow, constrained by numerous side conditions, and subject to many errors and aberations. A search for better algorithms resulted first in a comprehensive survey of image-warping techniques [40]).

The literature is largely silent on the problem of efficiently and smoothly mapping between two image regions which are delimited by arbitrary closed curves; such regions do not have the universally assumed four corners. The second result was the specification and verification of an algorithm that instead treats an image region as a collection of interior layers around a skeleton [41, 39]. These layers impose a type of local polar coordinate system which allows each shape to be "unwrapped" into a tree-like representation. Region-to-region warping is then defined by a natural mapping between the two resulting trees. Although there is no a priori way of defining quality of mapping, the results are aesthetically pleasing.

The third and most recent product is a new, highly efficient, general method for achieving 2-D image warps by separating the 2-D transform into two successive 1-D warps [42, 43]. It therefore extends the power of existing hardware systems that perform more limited classes of transformations by similar decompositions. However, this method shows that off-the-shelf hardware, in the form of digital filters with only minor modification for 1-D image resampling, can be used to realize arbitrary mapping functions cheaply and at video rates.

Further work will make use of this approach for performing high-speed elastic matching of deformed images. By using the spatial lookup tables introduced here, improved metrics for the quantification of deformation are possible. Extensions to 3-D may also be straightforward.

1.3 Optic Flow, and Rotational Motion

Optic flow computations are traditionally cast as continuous partial differential equations, but then are solved by discrete difference methods. Although there have been numerous approaches to the problem, differing in both equations and boundary conditions, few results have been obtained concerning the quality of solutions and their error. However, when the problem is cast in the domain of smoothing splines, and if boundary flow values obey Dirichlet constraints, several results are possible [30]. There is a unique solution; sparse, iterative methods can be sued to solve the resulting discrete system; error can be predicted. Further, the Chebyshev method of solution requires little global exchange of information, so it is eminently suited for parallelization. These results appear to be applicable to other low-level vision problems as well.

Looking now instead to the problems of smooth flow by a single object in three space, it is apparent that the understanding and interpolation of rotational motion (as in a "perfect spiral" football pass) is important in computer animation, robot control, and hypothesis-guided computer vision. A new, closed-form algorithm for doing so has been implemented, based on representing motions as quaternions on the unit three-sphere [31]. Resulting displays of interpolated values, and the computer animation sequences based on them, are smoother and more perceptually realistic than existing methods.

2 Middle-level Vision

We have exploited the theory of physical stereo to produce many methods for determining object position, orientation, and curvature. There are at least five now. We have found several powerful alternatives to standard regularization methods, and one of them has led to a non-traditional, one-step method for stereo matching, surface reconstruction, and segmentation. It forms the further basis for a novel stereo-texture fusion system, and holds the promise of further fusion work, possibly with depth-from-focus. The fusion work has already lead to an operational system in use outside the vision community. Research on shape-from-shadows has yielded a patent and optimal, parallelizable, distributed algorithms.

2.1 Physical Stereo

The theory of generalized physical stereo has produced five different applications at the middle levels of vision.

In the first, local surface orientation can be calculated by varying the wavelength and/or the linear polarization of a single incident light source [53]; only two settings of a polarizer are necessary for uniqueness of solution. This is a motionless variation on photometric stereo, and has been one of the earliest results from the theory. However, further study of the method has shown that the process of computing local surface normals can be made to rely simply on the empirical determination of the polarization Fresnel ratio; this is the same parameter necessary for determining the materials of an object and the components of a reflection. Thus, a formerly implicit factor is now seen to be a unifying parameter.

More practically, a second new technique to measure local surface orientation has been based on a more complete theory of the reflection of light. This theory combines the Torrance-Sparrow theory of reflection with the Wolff polarization theory of "quasi-monochromatic" (monochromatically filtered) light [46]. The technique enables surface orientation to be uniquely measured in arbitrary lighting by placing a simple monochrome filter and a linear polarizer in front of the sensor; two images taken at two orientations of the polarizer suffice. The equations that govern the calculations, called the polarization state matrix equations, are elaborate, but they are only a special case of the larger family of generalized physical stereo imaging equations.

These equations can be exploited to derive a third technique: the accurate determination of second order variations of smooth object surfaces as a function of height above the image plane. This technique uses a generalization of surface Hessian methods, which overconstrains the solution for the surface Hessian matrix, giving the second order variations of the smooth surface.

A fourth and very recent method for determining local surface orientation is based on a new imaging concept, the "photometric flow field" across an optic sensing array [47, 54, 55]. Conventional optic flow considers the rate of change in the physical position of the image of an object, as the object actually moves in three-space. In contrast, photometric flow considers the rate of change in the image irradiance of the image of a stationary object, as the illumination geometry moves in three-space instead. Such photometric flow fields can be used to determine local surface orientation and surface curvature. The method may be generalizable to extended light sources.

A fifth corollary to the theory of generalized physical stereo is a method to compute surface orientation from the stereo correspondence of linear features such as polygonal edges, or internal linear markings or texture [48, 56, 57]. It is in contrast to standard stereo, which uses point correspondences to compute the orientation of a plane from the 3-D position of three or more coplanar points. Stereo using line correspondence instead computes the orientation of a plane from the orientations of two or more coplanar lines. In the ideal world these two methods are exactly equivalent. But in the experimental world with measurement error, the errors inherent to measurement of surface orientation from line correspondence stereo does not grow nearly as fast with respect to baseline translation errors or with respect to distance from the baseline. Analysis and Monte Carlo simulations are shown to support this. There may be other vision algorithms which use also profit from the use of equivalent geometric constructions to combat error.

2.2 Regularized Surface Reconstruction

Defining the meaning of "smooth surface" is one of the burdens of surface regularization. In a survey paper, some of the benefits promised by the regularization framework are contrasted to some of its unheralded difficulties, particularly the problems of determining appropriate functional classes, norms, and regularization stabilizing functionals [8]. When regularization is subjectively tested via established procedures of psychology, the results of the methodology applied to the surface reconstruction problem often gives worse results than certain other non-traditional formulations (which are also presented and analyzed).

One of these non-traditional methods provide the basis for a non-heuristic algorithm which simultaneously reconstructs surfaces and segments the underlying data according to the same energy-based smoothness measure [9]. It is founded on the use of reproducing kernel-based splines, which allow efficient calculation of upper and lower bounds on surface energy. The system naturally deals with occluded objects, and also with sharply slanted surfaces, such as roads as seen from a vehicle.

This work on non-heuristic segmentation has been further extended into the development and testing of a new unified approach to stereopsis; it identifies the stereo matching criteria with the already combined non-heuristic reconstruction and segmentation criteria [13]. This energy criterion can be interpreted as a measure of match ambiguity, which is used to rank order all potential stereo matches. Stereo matching, surface reconstruction, and surface segmentation are therefore done in one step, according to one criterion. In tests so far, the method results in fewer unmatchable features than the Marr-Poggio-Grimson method. A parallel implementation is planned, to be followed by comparative performance analyses under various formulations of surface energy, and for various scenes.

2.3 Sensory Fusion

Existing work on the fusion of five different shape-from-texture methods has suggested a novel approach for classifying textures [28]. Each of the methods is tuned to certain image phenomena; the five are shape from spacing, shape from orientation, shape from size, and shape from absolute and relative eccentricity. Given a single texture patch, particularly one under perspective, each method will respond differentially according to the degree it believes the patch possesses cues that the method can exploit to derive shape information. These differential strengths can now all be gathered together as a signature feature vector for the texture. Although such vectors may not have any easily assignable "natural meaning", they can be manipulated in the usual way by standard pattern recognition or image segmentation techniques.

Having found ways of integrating into one process the three steps of stereo perception, and into another process five methods of texture perception, it was inevitable that the two processes themselves would be fused [29]. The resulting system now combines information in two fundamentally different ways, by intra-process and inter-process integration. For standardization reasons, inter-process integration necessarily incorporates a priori assumptions about surfaces, such as degrees and measures of smoothness; it communicates such data in a standardized way via a blackboard organization. In operation, the stereo process uses the relative accuracy and sparseness of the centroid of texels to begin feature localization, later switching to traditional zero-crossings. The work is further characterized by the choice of smoothness measure; roughly it minimizes variation in the 1.5 derivative, not the second. Final integration is achieved by weighting the surface constraints that are output by a process, by an amount that is inversely proportionally to the peak number of constraints a process can output; otherwise stereo, which is denser, would always outrank texture processing.

Applying this fusion technology to a real-world problem led to the successful completion of an operational system for oceanographers. These programs, now in constant use by researchers mapping strctures beneath the ocean floor, integrate navigational and positional information in order to recover the path of smoothly moving ocean vessels. The system's use of smoothing splines is backed by a clever heuristic to ignore faulty outliers in the data. The analysis and review of the project includes documentation of the negative results produced by more standard, "optimal" methods [10]).

Further pursuing the idea of multi-sensor fusion, initial re-implementation and testing has begun on algorithms for depth-from-focus. The experimental project will implement the three leading depth-from-focus algorithms, in order to comparatively determine their cost/accuracy trade-offs. The most efficient one becomes a candidate for further sensor fusion studies.

2.4 Shape from Dynamic Shadowing

The discrete version of a method for extracting surface shape information based from object self-shadowing under moving light sources has been awarded its patent [26].

The continuous version has seen extensive analysis, leading to a optimal, parallelizable algorithm [20, 21]. The two-dimensional problem is solved by decomposing it into a series of one-dimensional slices in the plane of the moving light source; these can be solved in parallel. Each strip is computed using as a basis a family of interpolating splines of an unusual piecewise linear form. The solution is checked against a side system of inequalities in order to preserve the implicit information that points interior to a shadowed region must lie below that shadow line; if the solution fails, a non-linear approximation algorithm accommodates the failing constraints.

The problem has a natural parallelization, not only into slices, but also into hill-and-valley segments; the latter parallelism has been implemented on a loosely coupled network of workstations. A smoothing spline approach has been developed to regularize noisy data. The question of optimal information (i.e., where to put the illuminants) has been solved in some very restricted cases; basically, the problem is dominated by the tangent of the incoming light ray angle. A full analysis of optimal light placement is being pursued.

3 Spatial Relations

We have invented, explored, or improved several representation schemes for objects they occupy and the light they reflect or obscure: lines, polyhedra, superquadrics, generalized cylinders, and sensor models. We have (literally) surveyed the representations of empty space, and how the representations can be efficiently changed as objects move. Even in one dimension, navigation is provably hard; we are examining two and more, in simulation and in the lab.

3.1 Representations of Objects

A new representation for a line in Euclidean three-space has been discovered, which uses only four parameters, the minimal number allowable, and still avoids singularities and special cases [32]. Without sacrificing convenience of computation, it is no longer necessary to represent lines in the more traditional six-parameter forms (such as Plucker coordinates, or point-and-orientation form), although the new representation has the added advantage that it is easy to convert to those forms. The representation, involving two parameters for position and two for orientation, readily generalizes to Euclidean n-space, where it uses 2n-2 parameters: n-1 for position, and n-1 for orientation.

When modeling objects by means of superquadrics, the primary concern in parameter estimation is the proper choice of the error-of-fit measures that control the nonlinear least square minimization techniques. The effectiveness of four such measures was tested on many examples using noisy synthetic data and actual range images, including multiple views of the same object, and including a superellipsoid with negative volume--the latter being an important primitive for constructive solid geometry-based modeling. Existing measures of fit appear inadequate, and a new one that performs significantly better was developed and verified [17]. In process is the verification of these predicted differences in complete recovery systems using real data.

A related model of volumes, generalized cylinders, is not nearly as well-defined as superquadrics are. Only certain subclasses appear to be well-specified and well-behaved under reflectance. It would be valuable to be able to quickly and cheaply test an image for the presence of a member of one of these subclasses; these tests could serve as gatekeepers to more expensive algorithms in a general polymorphic shape recovery system. The test need not calculate any parameters; it might exploit invariants that simple confirm or deny membership without any attempt at reconstruction. One such subclass, the straight homogeneous generalized cylinders, can be shown to possess a limited form of such invariants, under various rotational transformations and imaging conditions [18]. The test make good use of contour information as well as image intensity; contour is most useful in recovering the axis, and intensity in recovering any tilt. A prototype system is under construction.

Another new project, the PROVER System (Parametric Representation of Volumes: Experimental Recovery System) is designed to allow numerical recovery of parametric representations from multiple types of data, and multiple sensor types. An important feature of the system is its use of explicit sensor error models. Initial implementations are underway, and a prototype system with restricted parametric representations and data types is already running. The system will be used to develop accurate sensor error models, and will help demonstrate the effect of such models in the recovery of parametric volumes. Because of the significant computation cost of the approach, a parallel implementation is already underway.

Experience with merging multiple sensor data sources usually results in examining the sensor modeling problem from the perspective of the automatic generation of viewpoint, geometric, and sensing constraints. Assuming an assembly or an inspection domain, such an analysis is based on both CAD/CAM object models and low-level sensor models. The emphasis is on the automatic and intelligent handling of partial object descriptions, and partial or total sensor occlusions. The automatic generation of sensor viewpoint is a natural place to begin. The goal is to be able to automatically select a viewpoint for a vision sensor from which features of interest on an object will satisfy particular constraints in the image, among them, visibility. A prototype system has been developed that computes the regions in space where a face of an object occludes the target features [38]. The geometric model of the object is polyhedral, but its faces may be concave and multiply-connected.

3.2 Representations of Space

A survey of some 80 papers dealing with environmental representations of mobile robots has been completed and revised [19]. Most of these representations assume a static two-dimensional world, and a complete bird's-eye knowledge of free space and obstacles. The survey also proposes a taxonomy of this new field: it describes map primitives, such as frames of reference and map symbols, and representations, such as dehydrated free space (mixed polyhedra, and vertex graphs), simple mosaics (tessellations, distance maps, and quadtrees), and reconstituted free space (convex cells, and freeways). There continues to be a relative paucity of results on qualitative, topological navigation, however.

Extending previous work on path planning in dynamic environments using digital distance maps [11], complexity bounds have recently been derived on the constrained distance transform for computing digital distance maps. Further, the method has been extended to handle path planning with spatially varying distance metrics. In particular, digital terrain maps (currently synthetic) can provide auxiliary information (for example, surface height and ground-cover) that affects distance measures in a spatially-varying way. Such spatially varying distance cost problems are relatively frequent, and vertex based algorithms do not generalize well to these problems; their strengths under dynamic updating, however, are being investigated.

3.3 Theory and Practice of Navigation

A model has been formalized for topological navigation in one-dimensional spaces, such as along single roads, corridors, or transportation routes; it demonstrates that the problem is surprisingly difficult computationally [24, 25]. The model includes three levels of abstraction: the concepts and representations of the world itself (a version of "Lineland"), the world as abstracted into symbols and landmarks by an omniscient map-maker, and the world as experienced by a limited navigator who follows the map-makers directions. Having also modeled the navigator's sensors in a primitive way (a sensor here being more like a feature detector), it is straightforward to show that the problem of choosing an effective and efficient subset of sensors for navigation via landmarks is NP-complete. However, simplifying heuristic evaluation functions do exist, and are being explored for their effectiveness. The method has also been extended to a grid-like version of two dimensions, with similar results. It still remains that a "good" set of directions is ill-defined and intractable.

Work on the mobile robot platform of AT&T Bell Laboratories continues; sonar and custom VLSI vertical-edge detecting vision now cooperate, albeit weakly. The edge-tracking Kalman filter has been further refined, and initial models of the corridors and their effect on vertical edge positioning is being investigated.

4 Parallel Algorithms

We have analyzed the performance of the parallelization of several computationally optimal algorithms for depth interpolation; since the problem is typical of others at the low-level of vision, the optimality results should easily transfer. We have invented a particularly simple, accurate, and robust shape-from-texture algorithm based on image autocorrelation that appears to outperform human observation on real scenes of roads, dirt, and grass. We have designed and implemented a near-optimal programming environment for validating parallel pyramid-based SIMD algorithms on the Connection Machine. On our PIPE, we are implementing a system for optic flow determination that fuses the results of intensity correlation methods and spatiotemporal energy methods; the method has already generated a robust grey-level corner detector as an offshoot. The PIPE is fast enough to provide real-time robot arm control information, which we are preparing to demonstrate by the dynamic grasping of moving objects.

4.1 SIMD Algorithms

Many constraint propagation problems in early vision, including depth interpolation, can be cast as solving a large system of linear equations where the resulting matrix is symmetric, positive definite, and sparse. Analysis and simulation of several numerical analytic solutions to these equations for a fine grained SIMD machine with local and global communication networks (e.g., the Connection Machine) shows that two methods are provably optimal in terms of computational complexity [14, 15, 16]). For a variety of synthetic and real range data, the adaptive Chebyshev acceleration method executes faster than the conjugate gradient method, if near-optimal values for the minimum and maximum eigenvalues of the iteration matrix are available.

When these iterative methods are implemented in a pyramidal multigrid (coarse-medium-fine) fashion, using a fixed multilevel coordination strategy, the multigrid adaptive Chebyshev acceleration method executed faster than the multigrid conjugate gradient method again. This appears to be the case because an optimal Chebyshev acceleration method requires local computations only. These methods have now been validate on actual range data.

As a possible front-end to such depth interpolation tasks, a new method for determining local surface orientation was developed from rotationally invariant textures based on the two-dimensional two-point autocorrelation of an image [12]. This method is computationally simple and easily parallelizable, uses information from all parts of the image, assumes only texture isotropy, and requires neither texels nor edges in the texture. Applied to locally planar patches of real textures such as roads, dirt, and grass, the results are highly accurate, even in cases where human perception is so difficult that people must be assisted by the presence of an artificially embedded circular object. However, follow-up extensions attempting to use the method for non-isotropic textures, even with built-in heuristic biases, were not successful. Nevertheless, the algorithm has several exploitable mathematical elegancies, and is amenable to parallel implementation.

As part of our efforts under Strategic Computing, three programming environments that support research on stereo and texture algorithms were developed, in parallel image pyramid style [22]). The current and final programming environment has been designed, installed, and documented; it is a highly efficient pyramid machine emulator that executes those image function primitives on the (University of Syracuse) Connection Machine 2. It cleverly reduces communication contention by an elegant, and probably optimal, embedding of the pyramid within the hypercube network. Mesh operations take only a small fixed amount of overhead proportional to the size of the hypercube; parent/child operations run in a smaller fixed time independent of hypercube size. This code is publicly available.

4.2 Pipeline Algorithms

Real-time "pixel-parallel" versions of a variety of image processing algorithms have now been developed for our PIPE architecture. Based on our past experience with pipelined processors [33], already installed have been algorithms for spatial filtering, spatiotemporal filtering, and pyramid-based spatial processing. Most recently, a novel grey-level template-driven corner detector that combines the advantages of two previously orthogonal approaches has been designed and validated; it executes in real time [34].

One application of these real-time algorithms is in real-time motion tracking [1, 4]. The motion in a scene is found by using spatio-temporal filters on a PIPE. The PIPE is able to update motion energy centroids at 10 HZ and this information is used to update the position of an arm mounted camera which tries to keep the object centered in the field of view. Latencies in the communication system between arm and camera effectively reduce the arm movement rate to 4 HZ. The system is being developed in order to pick moving objects in real-time with our Utah-MIT hand.

Robustness of robotic algorithms is a paramount concern; such reliability can be achieved using an information-fusion based approach. A prototype system is under development that combines multiple cues for a visual measurement, along with an associated confidence; the grey-level corner detector is the first example. Next under investigation is image-flow extraction, using a unified mathematical framework for matching-based and gradient-based techniques [35, 36, 37]). The two techniques are nicely complementary; intensity correlation methods work best in structured scenes, and spatiotemporal energy methods are more suited for textured scenes.

5 Robotics and Tactile Sensing

We have made great progress in integrating a Utah-MIT hand into our robotics testbed. We have developed a number of low-level sensing and actuation primitives that allow one to easily program the hand for simple tasks. In addition, we have been exploring human psychology to understand the ways that humans use active touch and to apply these strategies to our robotics environment.

5.1 Integrated Environments

The Utah/MIT dextrous hand provides a new set of tools to study intelligent touch and grasping. Cartesian-based low level control algorithms for the hand, and a more hybrid scheme using both tendon force and tactile contacts will eventually be part of a comprehensive grasping environment. It will be capable of performing tasks such as locating moving objects and picking them up, manipulating man-made objects such as tools, and recognizing unknown objects through touch. In addition, the integrated programming environment will allow grasping primitives to be included in an overall robotic control and programming system that includes dextrous hands, vision sensors, and multiple degree of freedom manipulators [2, 5, 6].

The system has been used to perform a number of grasping tasks, including pick and place operations, extraction of circuit boards from card cages, pouring of liquids from pitchers, and removing light bulbs from sockets. These tasks have been programmed using DIAL, a parallel, graphical animation language developed by Steven Feiner. DIAL permits task-level scripts which can then be bound to particular sensors, actuators, and methods for accomplishing a generic grasping or manipulation task. We are currently exploring ways to extend an environment such as DIAL to allow programming of a hand to be a first-class primitive in a robotic programming environment.

5.2 Multi-fingered Object Recognition

It requires intelligence and model building to emulate the human ability to recognize objects haptically: that is, by only using external tactile sensors, and internal force and position sensors. However, superquadric models have proven to be surprisingly easy to recover from sparse and noisy sensor data [3, 7]. This appears to be because of their small number of parameters, and consequently their ability o recover the shape descriptions of a very large class of objects. Generic or prototypical recognition strategies are straightforwardly possible.

In experiments, a database of 6 objects consisting of undeformed superquadrics (a block, a large cylinder, a small cylinder) and deformed superquadrics (a light bulb, a funnel, a triangular wedge) was each recovered accurately, with extremely sparse data, typically 30-100 points. This is about 100 times less data than with range sensing, but it has the advantage of not being restricted to a viewpoint that only exposes half the object's surfaces to the sensor. Work is underway to extend this system to include segmented objects, multiple representations of objects, and the dynamic updating of representations.

Using piezo-resistive tactile sensors mounted on the Utah-MIT hand,, we are currently implementing robotic analogs of human haptic shape recovery methods such as shape from enclosure, shape from contour and shape from lateral extent.

6 References

1. Allen, P.K. Real-Time Motion Detection on a Frame Rate Processor. Extended Abstracts of the 41st Annual SPSE Conference, May, 1988.

2. Allen, P.K. "Integrating Vision and Touch for Object Recognition Tasks". *International Journal of Robotics Research 7*, 6 (1988).

3. Allen, P.K. 3-D Modeling for Robotic Tactile Object Recognition. Third International Conference on Robotics and Factories of the Future, August, 1988.

4. Allen, P.K. Real-Time Motion Tracking using Spatio-Temporal Filters. Proceedings of the DARPA Image Understanding Workshop, May, 1989. (These proceedings).

5. Allen, P.K., Michelman, P., and Roberts, K. "An Intelligent Grasping System". *IEEE Computer 22*, 3 (March 1989).

6. Allen, P.K., Michelman, P. and Roberts, K. An Integrated System for Dextrous Manipulation. IEEE International Conference on Robotics and Automation, May, 1989.

7. Allen, P.K., and Roberts, K. Haptic Recognition Using a Dextrous Multi-Fingered Hand. IEEE International Conference on Robotics and Automation, May, 1989.

8. Boult, T.E. Regularization: Problems and Promises. Extended Abstracts of the 41st Annual SPSE Conference, May, 1988.

9. Boult, T.E. and Liang-Hua Chen. Synergistic Smooth Surface Stereo. Proceedings of the International Conference on Computer Vision, December, 1988.

10. Boult, T.E., and Allen, B. Integration of Navigational and Positional Information to Recover the Path of a Smoothly Moving Vessel. Proceedings of the SPIE Conference Sensor Fusion, SPIE, November, 1988.

11. Boult, T.E. "Dynamic Digital Distance Maps". *IEEE Journal of Robotics and Automation* (To appear 1989), .

12. Brown, L.G., and Shvaytser, H. Surface Orientation from Projective Foreshortening of Isotropic Texture Autocorrelation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1988.

13. Chen, L.-H., and Boult, T.E. Analysis of Two New Stereo Matching Algorithms. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1988.

14. Choi, D.J. *Solving the Depth Interpolation Problem on a Parallel Architecture with Efficient Numerical Methods*. Ph.D. Th., Department of Computer Science, Columbia University, 1988.

15. Choi, D.J., and Kender, J.R. Solving the Depth Interpolation Problem on a Parallel Architecture with a Multigrid Approach. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1988.

16. Choi, D.J., and Kender, J.R. "Solving the Depth Interpolation Problem on a Parallel Architecture with Efficient Numerical Methods". *IEEE Transactions on PAMI* (In revision 1989).

17. Gross, A.D., and Boult, T.E. Error of Fit Measures for Recovering Parametric Solids. Proceedings of the International Conference on Computer Vision, December, 1988.

18. Gross, A.D. Straight Homogeneous Generalized Cylinders: Analysis of Reflectance Properties and a Necessary Condition for Class Membership. Proceedings of the DARPA Image Understanding Workshop, May, 1989. (These proceedings).

19. Hanvey, M. "Environmental Representations for Mobile Robot Navigation". *ACM Computing Surveys* (In revision 1989).

20. Hatzitheodorou, M., and Kender, J.R. An Optimal Algorithm for the Derivation of Shape from Shadows. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1988.

21. Hatzitheodorou, M. The Derivation of 3-D Surface Shape from Shadows. Proceedings of the DARPA Image Understanding Workshop, May, 1989. (These proceedings).

22. Ibrahim, H.A.H., Brown, L., and Kender, J.R. Parallel Vision Algorithms--Final Report. Tech. Rept. CUCS-415-89, Department of Computer Science, Columbia University, 1989.

23. Kender, J.R. Some Issues for the Panel on Methodology and Standards in CVPR Research. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1989.

24. Kender, J.R., and Leff, A. "Why Direction-Giving is Hard: The Complexity of Linear Navigation by Landmarks". *IEEE Transactions on Systems, Man, and Cybernetics* (To appear 1989).

25. Kender, J.R., and Leff, A. Why Direction-Giving is Hard: The Complexity of Linear Navigation by Landmarks. Proceedings of the AAAI Spring Symposium Series, March, 1989.

26. Kender, J.R., and Smith, E.M. A Method and an Apparatus for Determining Surface Shape Utilizing Object Self-Shadowing. Patent Number 4,792,696. December, 1988.

27. Proceedings of the IEEE. *Special Issue on Computer Vision*, August, 1988. editors Li, H., and Kender, J.R..

28. Moerdler, M.L. Multiple Shape from Texture into Texture Analysis and Surface Segmentation. Proceedings of the International Conference on Computer Vision, December, 1988.

29. Moerdler, M.L., and Boult, T.E. The Integration of Information from Stereo and Multiple Shape-from-Texture. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1988.

30. Lee, D., Papageorgiou, A., and Wasilkowski, G. Computational Aspects of Determining Optic Flow. Proceedings of the International Conference on Computer Vision, December, 1988.

31. Roberts, K.S., Ganapathy, S.K., and Bishop, G. Smooth Interpolation of Rotational Motions. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1988.

32. Roberts, K.S. A New Representation for a Line. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1988.

33. Singh, A., and Lala, P.K. "A Multilayer Cellular Architecture for a Highly Parallel VLSI Supercomputer". *IEEE Transactions on Computers* (To appear 1989).

34. Singh, A., and Shneier, M. Grey Level Corner Detection: A Generalization and a Robust Real Time Implementation. Tech. Rept. TR-89-099, Philips Laboratories, Briarcliff Manor, NY, 1989.

35. Singh, A. Information-Fusion: An Approach to Robust Image-Flow. Proceedings of the DARPA Image Understanding Workshop, May, 1989. (These proceedings).

36. Singh, A. Image-Flow Extraction: Information from Discontinuities. Tech. Rept. TR-89-017, Philips Laboratories, Briarcliff Manor, NY, 1989.

37. Singh, A. Aperture Problem: A Review and a New Local Solution. Tech. Rept. TR-89-082, Philips Laboratories, Briarcliff Manor, NY, 1989.

38. Tarabanis, Konstantinos and Tsai, Roger Y. Viewpoint Planning: The Visibility Constraint. Proceedings of the DARPA Image Understanding Workshop, May, 1989. (These proceedings).

39. Wolberg, G. Image Warping Among Arbitrary Planar Shapes. Proceedings of the Computer Graphics International Conference, Geneva, May, 1988. also appears in New Trends in Computer Graphics, editors N. Magnenat-Thalmann and D. Thalmann, Springer-Verlag, pp. 209-218, 1988.

40. Wolberg, G. Geometric Transformation Techniques for Digital Images: A survey. Tech. Rept. CUCS-390-88, Department of Computer Science, Columbia University, 1989. to be published by the IEEE Computer Society Press.

41. Wolberg, G. "A Skeleton-Based Image Warping". *Visual Computer* (1989). to appear.

42. Wolberg, G., and Boult, T.E. Separable Image Warping With Spatial Lookup Tables. Proceedings of the DARPA Image Understanding Workshop, May, 1989. (These proceedings).

43. Wolberg, G. and T. Boult. Separable Image Warping. Computer Graphics (SIGGRAPH 89 Proceeedings), , 1989. (To appear).

44. Wolff, L.B. Segmentation of Specular Highlights From Object Surfaces. Proceedings of the SPIE Optics, Illumination, and Image Sensing for Machine Vision III, 1988, pp. 198-205.

45. Wolff, L.B. Classification of Material Surfaces Using the Polarization of Specular Highlights. Proceedings of the SPIE Optics, Illumination, and Image Sensing for Machine Vision III, 1988, pp. 206-213.

46. Wolff, L.B. Accurate Measurement of Second Order Variations of a Smooth Object Surface. Proceedings of the SPIE Sensor Fusion: Spatial Reasoning/ Scene Interpretation, 1988.

47. Wolff, L.B. Shape From Lambertian Photometric Flow Fields. Proceedings of the SPIE Optics, Illumination, and Image Sensing for Machine Vision III, 1988, pp. 255-262.

48. Wolff, L.B. Measuring the Orientation of Lines and Surfaces Using Translation Invariant Stereo. Proceedings of the SPIE Sensor Fusion: Spatial Reasoning/ Scene Interpretation, 1988.

49. Wolff, L.B. Using Polarization To Separate Reflection Components. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1989. (To appear).

50. Wolff, L.B. "Classification of Material Surfaces using the Polarization of Specular Highlights". *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI* (In revision 1989).

51. Wolff, L.B. Material Classification and Separation of Reflection Components From Polarization/Radiometric Information. Proceedings of the DARPA Image Understanding Workshop, May, 1989. (These proceedings).

52. Wolff, L.B., and Boult, T.E. Polarization/Radiometric Based Material Classification. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1989. (To appear).

53. Wolff, L.B. "Spectral and Polarization Stereo Methods Using a Single Light Source". *International Journal of Computer Vision* (In revision 1989).

54. Wolff, L.B. Shape Understanding From Lambertian Photometric Flow Fields. Proceedings of the DARPA Image Understanding Workshop, May, 1989. (These proceedings).

55. Wolff, L.B. Shape Understanding From Lambertian Photometric Flow Fields. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1989. (To appear).

56. Wolff, L.B. Accurate Measurement of Orientation From Stereo Using Line Correspondence. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1989. (To appear).

57. Wolff, L.B., and Boult, T.E. Experiments For Determining Surface Orientation From Line Correspondence Stereo. Proceedings of the International Joint Conference on Artificial Intelligence, August, 1989.