CUCS-332-88

Appears in DARPA
Image Understanding
Workshop, 1988

Image Understanding and Robotics Research at Columbia University

John R. Kender[1]
Peter K. Allen
Terrance E. Boult
Hussein A. H. Ibrahim

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

0 Introduction

The research investigations of the Vision/Robotics Laboratory at Columbia University reflect the diversity of interests of its four faculty members, two staff programmers, and 15 Ph.D. students. Several of the projects involve either a visiting computer science post-doc, other faculty members in the department or the university, or researchers at AT&T Bell Laboratories or Philips Laboratories. We list below a summary of our interest 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.

0.1 Low-level Vision

0.1.1 Theories Involving Stereo

1. A unified theory of generalized stereo vision (Larry Wolff [45, 49, 50]).
2. The derivation of shape from polarizing surfaces (Larry Wolff [46, 47, 48]).
3. Optimal estimators for stereo triangulation error (Ken Roberts, Dr. S. Kicha Ganapathy of AT&T Bell Laboratories [34]).

0.1.2 Data Representations

1. A new representation for a line in three-space (Ken Roberts [35]).
2. Smooth interpolation of rotational motions (Ken Roberts, Drs. S. Kicha Ganapathy and Garry Bishop of AT&T Bell Laboratories [36]).

0.1.3 Applications to Graphics

1. Realistic rendering of scenes using polarization properties (Larry Wolff, Dave Kurlander [51]).
2. A new data structure and algorithm for the mapping of arbitrary shapes (George Wolberg [43, 44]).

0.2 Middle-level Vision

0.2.1 Regularized Surface Reconstruction and Stereo

1. A critical study of regularization methodology (Terry Boult [4, 10]).
2. Regularized surface reconstruction and segmentation based on smoothness energy (Terry Boult, Liang-Hua Chen [11]).
3. Integrated stereo matching, surface reconstruction, and surface segmentation (Terry Boult, Liang-Hua Chen [12, 16, 17]).

0.2.2 Sensory Fusion

1. Fusion of multiple shape-from-texture methods (Mark Moerdler, John Kender [30, 31]).
2. Fusion of texture and stereo (Mark Moerdler, Terry Boult [5, 32, 33]).

0.2.3 Shape from Dynamic Shadowing

1. A discrete method for deriving surfaces from dynamic shadows (John Kender, Earl Smith [26, 29]).
2. An optimal algorithm for shape from continuous shadows (Michalis Hatzitheodorou, John Kender [22, 23]).

0.2.4 Application to Range Data

1. Recovery of superquadric parameters (Terry Boult, Ari Gross [6, 7, 13]).
2. Spline-based recovery of smooth oceanographic positional information (Terry Boult, Dr. Barry Allen of Columbia University's Lamont-Doherty Geological Observatory [8]).

0.3 Spatial Relations

0.3.1 Representations of Objects and Space

1. Analysis and extension of issues in aspect graphs (John Kender, David Freudenstein on leave, Prof. Jonathan Gross [27]).
2. Survey of algorithms for the representation of space (Monnett Hanvey [21]).

[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.

3. Efficient updating of digital distance maps in dynamic environments(Terry Boult [9]).

0.3.2 Theory and Practice of Navigation

1. Landmark definition and the representation and complexity of custom maps (John Kender, Avraham Leff [28]).
2. Systems issues in practical real-time robotic navigation (Monnett Hanvey, Drs. Bob Lyons and Russ Andersson of AT&T Bell Laboratories).

0.4 Parallel Algorithms

0.4.1 Low- and Middle-level Vision Theory

1. Depth interpolation using optimal numerical analysis techniques on a pyramid machine (Dong Choi, John Kender [18, 19]).
2. Determination of surface orientation from foreshortened texture autocorrelations (Lisa Brown, Dr. Haim Shvaytser of Weizmann fellowship [14, 15]).

0.4.2 Research and Applications on Tree Machines

1. Simulators and programming environments for Non-Von and for the Connection Machine (Hussein Ibrahim, Lisa Brown [20, 24, 25]).
2. Stereo, texture, and other pyramid-based algorithms (Hussein Ibrahim, Lisa Brown).

0.4.3 Research and Applications on Pipelined Machines

1. Implementing basic real-time image algorithms for pipelined processors (Ajit Singh, Peter Allen [37, 38, 41]).
2. Sensor fusion of correlation and of spatio-temporal approaches to optic flow (Ajit Singh, Peter Allen, Dr. Surendra Ranganath of Philips Laboratories [39, 40, 42]).
3. Real-time object tracking and interception algorithms (Peter Allen [2]).

0.5 Robotics and Tactile Sensing

0.5.1 System Development

1. Cartesian-based control of the newly-acquired Utah hand (Ken Roberts, Peter Allen).
2. Interfacing proprietary skin-like tactile sensors (Peter Allen).

0.5.2 Multi-fingered Object Recognition

1. Sensor models and CAD/CAM object models (Peter Allen, Dino Tarabanis [1, 3]).
2. Haptic recognition via active exploration with a instrumented robot hand (Ken Roberts, Peter Allen)

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 extended our work on a generalized framework for the perception of physical surface properties, and have formalized the conditions under which image measurements can break certain symmetries of observation in order to uniquely define depth-related object values. We have applied this general theory to the specific case of the derivation of surface orientation from differences in the polarization of reflected light, and have shown that only two settings of a linear polarizer placed before the camera are necessary for uniqueness. In other work, we have analyzed the error in traditional two-camera parallax stereo using a Bayesian statistical approach, and have computed optimal estimators that are extensible to multi-camera imaging configurations.

Some of our work has lead to economies of data representation; in particular, we have discovered a new way of representing lines in three-space that requires only four parameters and is totally free of annoying special cases. In work on rotational motions, we have defined an efficient, closed form way of interpolating their representations as quaternions over the associated three-sphere; the method leads to surprisingly smooth animations.

Work on low-level vision often leads to corresponding results in graphics. We have empirically validated that our theory of polarization adds striking realism to the computer graphic generation of certain types of scenes involving reflections. Lastly, in the course of investigating efficient object tracking algorithms, we have devised and implemented a general but fast method for mapping arbitrary planar shapes onto each other, based on a new skeletonization data structure.

1.1 Theories Involving Stereo

1.1.1 Generalized Stereo Vision

Generalized stereo begins as an abstract unification of two distinct exiting stereo techniques: traditional parallax stereo, which calculates surface depth by varying the camera focal point, and photometric stereo, which calculates surface orientation by varying the light positions. A generalized stereo method calculates arbitrary visual object features (world coordinate position, local surface orientation, Gaussian curvature, color reflectivity, etc.) by varying related physical imaging parameters (position of focal point, orientation of incident light source, polarization of incident light source, etc.) The object feature is determined by the intersection, in a parameter feature space, of solution loci generated from a system of equations relating features to parameters [45, 49, 50].

We have shown that in its formalized axiomatic definition, a generalized stereo method is characterized by four things: a visual object feature to be measured, a functional way of converting image observables such as image intensity into other observables such as image gradient, a set of variable imaging parameters, and the equations relating all three. We have illustrated the theory with many examples.

Additionally, we have characterized the error intrinsic to this family of methods by noting that the dimension of measurement ambiguity is readily determined by the implicit function theorem applied to the equations at the point of intersection. More accurately, error can be characterized in terms of symmetries of solution loci using the theory of groups. We have established two theorems which state the precise conditions under which the intersection of solution loci can be further disambiguated.

1.1.2 Polarization Stereo

We have investigated several applications of the generalized stereo theory. We have analyzed how surface orientation can be calculated by varying the wavelength and/or linear polarization of a single incident light source [46]. More practically, we have also proposed a new technique to measure local surface orientation based on a more complete theory of reflection of light. This theory combines the Torrance-Sparrow theory of reflection with the Wolf polarization theory of "quasi-monochromatic" (monochromatically filtered) light [47, 48]. 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 stereo imaging equations.

1.1.3 Optimal Stereo Triangulation Techniques

We have analyzed the positional error in stereo triangulation using a Bayesian statistical approach, and have derived optimal estimators based on several different sets of imaging assumptions [34]. One assumption models the camera error function in a new and more general way, by including a depth-sensitive (1/z^n) factor. Our techniques are elegantly extended to the case of more than two cameras.

Intuitively, we prove that the following methods are optimal. For a given stereo pair, reject any errors perpendicular to the epipolar line. Weight each camera's estimate of source point position by the reciprocal of the variance of its error function and the square of the depth of the source point from it. For more than two images, compute the results taking the images pairwise, then combine them by weighting each result by the square of the pair's baseline.

1.2 Data Representations

1.2.1 Representation of Three-Space Lines

We have constructed a new representation for a line in Euclidean three-space which uses only four parameters, the minimal number allowable, and still avoids singularities and special cases [35]. Therefore, 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.

1.2.2 Interpolation of Rotational Motion

Smooth interpolation of rotational motion (as in a "perfect spiral" football pass) is important in computer animation, robot control, and hypothesis-guided computer vision. We have implemented a new, closed form algorithm for doing so, based on representing motions as quaternions on the unit three-sphere [36]. Resulting displays of interpolated values, and the computer animation sequences based on them, are smoother and more perceptually realistic than two existing methods.

1.3 Applications to Graphics

1.3.1 Rendering Using Polarization Properties

We have applied the more complete theory of reflection developed in the work described above to ray tracing algorithms in computer graphics. We report striking differences in the rendering of certain scenes involving reflections when the phenomenon of polarization is included [51]; the differences are preferred by observers. This lends some evidence to the belief that the reflection model is at least qualitatively very correct.

1.3.2 Mapping of Arbitrary Planar Shapes

In attempting to analyze and track objects between images, we discovered that the literature was 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. We have specified and verified an algorithm that instead treats an image region as a collection of interior layers around a skeleton (similar to that in [43]). 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 [44]. Although there is no a priori way of defining quality of mapping, the results are esthetically pleasing.

2 Middle-level Vision

We have presented a critical overview of the regularization methodology, and have demonstrated new means of specifying the function class and its stabilizing functional that, although non-traditional, give qualitatively better results. We have exploited one of these ways, which is heavily dependent on the use of reproducing kernel-based splines, to surface segmentation; the method computes upper and lower bounds on local surface energy prior to surface labeling, and demonstrates good results on synthetic and real image and range data, and even on some transparent surfaces. Further, we have incorporated this energy-based approach into a system that integrates the formerly separate middle-level vision stages of stereo matching, surface reconstruction, and segmentation into a more straightforward one-step surface labeling based on a single measure of ambiguity; quantitatively, it results in a significantly higher percentage of correct matches.

We have designed, built, and verified on synthetic and real imagery, a blackboard-based system that fuses the independent and occasionally conflicting information from multiple (four or more) texture cues into a integrated method for surface segmentation and orientation determination; it is organized around a new image data structure, the augmented texel, and achieves sensor fusion via a Hough-like method on a trixelated Gaussian sphere. We have extended the method to a design and preliminary system (tested on a real image) that fuses the resulting surface orientation with the results of the one-step stereo method described above; this design thus coordinates the two intra-modality integrations with an inter-modality relaxation-based fusion of information through a weighted averaging, according to a non-traditional "smoothness norm", of zero-crossing and texel-centroid data.

Our work on the derivation of surface information from self-shadowing has resulted in a patent application for the discrete case. Additionally, we have analyzed the continuous case according to methods of functional analysis and have devised a provably optimal algorithm for surface recovery that is grounded in an unusual family of basis-like splines and an unusual iterative procedure for handling the non-linearity of the mutual illumination constraints; we have demonstrated a high degree of surface reconstructive accuracy on one-dimensional data.

Using a nonlinear least square minimization technique on the so-called inside-out function, we have designed and demonstrated a system for the robust recovery of superquadric parameters from both noisy synthetic and actual range data imagery, including even the case of a superellipsoid with negative volume (a construct used in solid modeling). Transferring our middle-level vision technology to a real world problem, we have begun to analyze various methods for inferring the geological structures below the surface of the ocean by first fusing several noisy sources of ship-board sensory data, such as satellite, dead reckoning, and gravitational information; this system for position tracking is now in regular field use.

2.1 Regularized Surface Reconstruction and Stereo

2.1.1 Critical Analysis of Regularization

We have presented a survey of some of the benefits promised by the regularization framework, and also of some of its difficulties, particularly the problems of determining appropriate functional classes, norms, and regularization stabilizing functionals [4]. When we subjectively tested (via established procedures of psychology) the results of the methodology applied to the surface reconstruction problem, we found that non-traditional formulations provided better results. It is not surprising that we were then able to document the lack of development of most of the promises of regularization theory, finding only three actual examples of its fruitful realization [10].

2.1.2 Energy-based Surface Segmentation

Although current surface reconstruction algorithms have strong foundations in mathematics, the segmentation aspects of the work are purely heuristic. We have developed and tested a non-heuristic algorithm which simultaneously reconstructs surfaces and segments the underlying data according to the same energy-based smoothness measure [11]. It is founded on the use of reproducing kernel-based splines, which allow efficient calculation of upper and lower bounds on the energy. The system naturally deals with occluded objects, and also with sharply slanted surfaces, such as roads as seen from a vehicle. We have verified the system on a gamut of artificial and natural data, including transparent surfaces.

2.1.3 One-step Stereo Matching, Reconstruction, and Segmentation

Traditional stereopsis is done in three phases: 1) suitable features are detected in each image, 2) corresponding features are matched and disparity is determined, and 3) a complete depth map is approximated and segmented. We have extended our work on non-heuristic segmentation by developing a new, one-step approach to stereopsis that unifies the stereo matching criteria with our already combined reconstruction and segmentation criteria [12, 16, 17]. The criteria is exploited in the form of a measure of match ambiguity, which is used to rank order all potential matches. The method results in fewer unmatchable features than the Marr-Poggio-Grimson method.

2.2 Sensory Fusion

2.2.1 Fusing Shape-from-Texture Methods

We continue to augment and refine our system for integrating various modalities for determining surface orientation from multiple, independent, conceptually parallel, and possibly conflicting textural cues. The system uses a new data structure, the augmented texel, which combines multiple constraints on orientation, each with its own assurity weighting, in a compact notation for a single surface patch [30, 31].

We have demonstrated the system using four texture modules (shape from spacing, eccentricity, orientation, and size), on both synthetic and real imagery of surfaces, some of them curved or transparent, with robust results: slant and tilt are usually recovered within a few degrees. We have shown examples of real surfaces for which individual texture methods fail to determine surface orientation accurately because of noise, but for which their fusion succeeds. Part of the noise tolerance of the system is derived from the relaxation refinement of initial hypotheses about surface orientation and extent, which themselves are derived from a (noise tolerant) Hough accumulation array on the surface of the trixelated Gaussian sphere.

2.2.2 Fusing Stereo with Texture

Having found ways of integrating into two separate processes the three steps of stereo perception and at least four methods of texture perception, we have combined our results in a single system that fuses stereo and texture together [5, 32, 33]. Although it is still under development (it has processed only a single real image), it is uniquely structured to provide two qualitatively different means of information fusion, namely, intra-process and inter-process integration. The latter incorporates a priori assumptions about surfaces, such as degrees and measures of smoothness, and communicates such data via a blackboard organization. Such a two-stage organization does not appear inconsistent with what is known about human visual modularization.

In particular, 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 done by weighting the significance of a surface constraint produced by either process inversely proportionally to the total number of constraints the process outputs (otherwise stereo would always outweigh texture processing).

2.3 Shape from Dynamic Shadowing

2.3.1 The Discrete Case: Shape from Darkness

We have analyzed and validated on synthetic data a new method, called shape from darkness, for extracting surface shape information based on object self-shadowing under moving light sources [26]. Unlike most shape-from methods, it does not require a reflectance map, and it works on non-smooth surfaces. Shadow information is stored in a novel data structure called the suntrace, which records the quantized angle of illumination at which a given image point was first illuminated. Given n points, the surface reconstruction problem becomes the satisfaction of 8n constraint equations in 2n unknowns, one unknown each for the upper and lower surface bound for each image point. An unusual form of relaxation, in which pixels can affect other pixels at a great distance, quickly converges to the solution. Columbia University has applied for a patent on the method [29].

2.3.2 The Continuous Case: Optimal Shape from Shadows

We have analyzed the same problem in the continuous setting, decomposing the two-dimensional problem into a series of one-dimensional slices in the plane of the moving light source. Casting the problem in a Hilbert space, we derived a provably optimal algorithm which involves interpolating splines of an unusual piecewise linear form [22, 23]. A side system of inequalities is optionally invoked in order to preserve the implicit information that points interior to a shadowed region must lie below that shadow line. The problem has a natural parallelization, not only into slices, but also into hill-and-valley segments. Our implementation has demonstrated high accuracy using few light sources on even badly nondifferentiable test functions. We are now attempting to analytically determine optimal light source placement.

2.4 Application to Range Data

2.4.1 Recovery of Superquadrics

Many have noted the simultaneous descriptiveness and compactness that superquadrics offer as a volumetric model; noted, too, is their well-defined inside-out functions needed in parameter recovery. However, we have determined that the primary concern in superquadric parameter estimation is the proper choice of the error-of-fit measures that control the nonlinear least square minimization techniques. We have explored the effectiveness of several such measures on many examples using noisy synthetic data and actual range images, including multiple views of the same object and a superellipsoid with negative volume, the latter being an important primitive for constructive solid geometry-based modeling. We have concluded that existing measures of fit are inadequate, and have proposed ones that perform better [6, 7, 13].

2.4.2 Recovery of Oceanographic Positional Information

We have investigated the problem of integrating different types of positional information, such as various satellite and inertial data, in order to reconstruct the path taken by an exploratory geological ocean vessel. Typical paths are piecewise very smooth except at turns; the problem is therefore a one-dimensional analogue of the middle-level vision problem of smooth surface recovery from sparse depth data [8]. We further investigated the related two-dimensional analogue problem: the inference of the geological structures below the surface of the ocean floor from gravitational information. The problem was solved by again using smoothing splines, backed by a clever heuristic to ignore faulty outliers in the data. The system, with some amount of human intervention, has been put into regular use by the researchers at Columbia's Lamont-Doherty Geological Observatory.

3 Spatial Relations

We have extended the semantics of an object representation, called the aspect graph, to the more realistic imaging environments having finite camera precision; we fixed several definitional problems in the process. Having surveyed 60 papers on representations of navigational space, we have taxonomized the main approaches to the problem: they use dehydrated free space, simple mosaics, or reconstituted free space representations. We extended the path-planning representation, called the digital distance map, to dynamic environments, and presented an efficient algorithm for its maintenance under object movement.

We defined a model of landmarks and map-making, and showed that the problem of providing a navigator with a list of directions is, even in the simplest case, an NP-complete problem; nonetheless, heuristics exist to help cut down search in creating "good" maps (defined as being either "short" or "easy"). We have been programming a mobile robot platform, attempting to have it navigate topologically via landmarks such as corridor intersections.

3.1 Representations of Objects and Space

3.1.1 Aspect Graphs and Degenerate Views

An aspect graph is a representation of the effect that viewing angle has on an object's observable features. While attempting to extend this concept to formations of objects, we discovered several inadequacies and errors in its current definition and use [27]. We demonstrated that the key concept of "characteristic view" is not well defined; in fact, it rarely is defined at all. The problem becomes more acute under finite camera resolution, where idealized aspect graphs become more like spatial maps: both nodes and arcs now have width. Given camera resolution and object size, we were able associate probabilities of observation to certain "degenerate views" of some simple objects.

Our most recent work has noticed a close connection between the aspect graph and the so-called first barycentric subdivision of standard graphs in graph theory; we are attempting to exploit this and other formal similarities.

3.1.2 Representation of (Un)Occupied Space

We have completed a survey of some 60 papers dealing with environmental representations of mobile robots. Most of them assume a static two-dimensional world, and a complete bird's-eye knowledge of free space and obstacles. We have given a taxonomy of map primitives, such as frames of reference and map symbols, and a taxonomy of representations, such dehydrated free space (mixed polyhedra, and vertex graphs), simple mosaics (tessellations, distance maps, and quadtrees), and reconstituted free space (convex cells, and freeways). We have also noted the relative paucity of results on qualitative, topological navigation via landmarks.

3.1.3 Dynamic Digital Distance Maps

A digital distance map contains in each of its cells information about the distance and/or direction to some pre-specified goal set; if the environment is static, it makes path planning trivial. We have developed an algorithm that extends the utility of these maps to dynamic environments, such as robotic assembly domains [9]. The algorithm is efficient, in that it only updates those cells of free space that are in the moving object's "shadow", where a shadow is defined according to a precise but tricky grammar. The algorithm is two-phase: shadow calculation followed by map update; if the robot can avoid shadows, this allows some parallelization of robot movement with map updating. We have observed speedups of 25 times over brute force update. We are now extending the method to higher dimensions, such as configuration spaces, and investigating the use of multi-resolution techniques.

3.2 Theory and Practice of Navigation

3.2.1 The Computational Complexity of Map-Making

We have formalized a model of topological navigation in one-dimensional spaces, such as along single roads, corridors, or transportation routes, and have shown that the problem is surprisingly difficult computationally [28]. In our model, we carefully discriminated between 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), we proved that the problem of choosing an effective and efficient subset of sensors for navigation via landmarks is NP-complete.

However, the A* search procedure does apply; we also gave a simplifying heuristic evaluation function ("most new eliminated objects") for use with it. Having selected the proper sensor subsets, Dijkstra's shortest path algorithm gives the optimal set of directions for the navigator, where we defined an optimal map to be one that minimizes length of directions, cost of sensing, or some combination.

3.2.2 Driving the AT&T Mobile Robot

In work jointly supported by AT&T Bell Laboratories, we are investigating several systems issues in navigation by using their mobile robot platform. As an early experiment in landmark recognition, we have programmed it to track walls with its sonar; the robot notices intersections and dead ends, which are potentially significant external environmental cues for self-positioning. In related work, we have tackled the problem of calculating ranges to visually perceived vertical edges by using a simplified Kalman filter. Since the error introduced by quantization and other factors is not gaussian, this filter produces accurate estimates only at selected points; however, these estimates can be strategically combined using triangulation to increase accuracy. We are testing this filtering/triangulation system on the robot, aiming for 60 Hz cycle time.

4 Parallel Algorithms

We have analyzed the performance of the parallelization of several computationally optimal algorithms for depth interpolation, and have found that on a wide variety of synthetic and real range data the adaptive Chebyshev is the most efficient, even when implemented in a multigrid fashion. We have invented a particularly simple, accurate, and robust shape-from-texture algorithm based on image autocorrelation that outperforms human observation on real scenes of roads, dirt, and grass.

In our Strategic Computing work, we have designed and implemented three related programming environments for validating parallel pyramid-based SIMD algorithms; the third one elegantly exploits the Connection Machine's hypercube network to efficiently emulate a library of image functions for a virtual pyramid machine, at a fixed low overhead. We have also implemented, on simulators or the emulator, parallel pyramid-based stereo, segmentation, and Hough algorithms, as well as our new autocorrelation texture algorithm.

On our PIPE, we have built up a basic library of pipelined low-level image processing functions. We are implementing a system for optic flow determination that fuses the results of intensity correlation methods and spatiotemporal energy methods; the latter is based on a novel image structure called the pyramid of oriented edges. 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 Low- and Middle-level Vision Theory

4.1.1 Optimal Depth Interpolation

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. We have analyzed and simulated several numerical analytic solutions to these equations for a fine grained SIMD machine with local and global communication networks (e.g., the Connection Machine); the methods are provably optimal in terms of computational complexity. We have established that 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 [18].

We then extended these iterative methods by implementing them in a pyrmidal multigrid (coarse-medium-fine) fashion. Again we showed that, when used with a fixed multilevel coordination strategy, the multigrid adaptive Chebyshev acceleration method executed faster than the multigrid conjugate gradient method [19]. Further, we demonstrated that because an optimal Chebyshev acceleration method requires local computations only, it in turn executes faster than either adaptive Chebyshev acceleration or conjugate gradient methods, both of which require global computations. We continue to validate these algorithms on Utah laser range data.

4.1.2 Shape from Texture Autocorrelation

We have developed a new method for determining local surface orientation from rotationally invariant textures based on the two-dimensional two-point autocorrelation of an image [14, 15]. 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; it is thus more widely applicable than the method of Witkin. We have demonstrated that when 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.

Along the way, we have proven that the algorithm has several exploitable mathematical elegancies. For example, the autocorrelation of an isotropic texture will always be entirely composed of concentric scaled elliptic iso-contours; this makes the extraction of slant and tilt values from ellipse parameters nearly trivial. Secondly, because the autocorrelation has such robust structure, it is easy to filter out from it spurious noise such as that commonly generated by the horizontal smearing of pixels in typical CCD cameras.

4.2 Research and Applications on Tree Machines

4.2.1 Simulators and Programming Environments

As part of our efforts under Strategic Computing, we have developed three programming environments that support our research on stereo and texture algorithms, in parallel image pyramid style [20]. Our first environment assisted work on the fine-grained tree-structured SIMD Non-Von supercomputer (now discontinued) [24], and it consisted of simulators of various grain sizes. As Non-Von began winding down, we constructed a second, more abstract environment of image function primitives for general pyramid machine vision work; this environment was necessarily a transitional one for the preservation of the prior work. For our third and current programming environment, we have designed, installed, and documented a highly efficient pyramid machine emulator that executes those image function primitives on the University of Syracuse Connection Machine [25].

This third environment cleverly reduces communication contention by an elegant 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. The image functions allow the user to create pyramid data structures, to load/unload various pyramid levels, to move data up/down, and to perform several operations such as convolution and hierarchical operators on the created data structures. Both our texture and stereo work will benefit from the multiresolution capabilities: texture algorithms will adjust to texel size, and stereo will use feature-matching based on hierarchical correlation. Most recently, we are upgrading our environment to run on the CM2.

4.2.2 Pyramid-based Stereo and Texture

Our main objective under Strategic Computing is to develop, implement, and integrate parallel multi-resolution stereo and texture algorithms for determining local surface orientation and depth, to be used by autonomous land vehicle navigation systems.

We have implemented on the Non-Von simulator a straightforward parallelization of a multi-resolution version of the Marr-Poggio stereo algorithm, which achieves some economies on the SIMD architecture by exploiting the eight-fold symmetries of digitized Laplacian of Gaussian masks. We are parallelizing our new autocorrelation-based shape-from-texture technique for the Connection Machine, where it becomes technically even more elegant. Using image shifts to compute a finite window of the autocorrelation, we can compute surface orientation for surface patches in constant time.

In service to both of the above algorithms, we have implemented two parallel texture-based image segmentation algorithms and tested them on ERIM ALV road data. The first algorithm uses micro-edge density counts in dynamically varying windows that attempt to track the road edge from image to image; success has been limited by the low texture resolution. A second algorithm is both parallel and pyramid based, and is more successful. It constructs up-down links in the multi-resolution pyramid between nodes on adjacent levels, according to similarities of spatial grey level dependence statistics. By top-down iterative refinement of these links, the pyramid and hence the image is segmented; basically this is a parallel form of region-splitting. Lastly, we have implemented the generalized Hough transform, including the parallel creation of the reference tables.

4.3 Research and Applications on Pipelined Machines

4.3.1 Real-time Algorithm Library

We are developing real-time "pixel-parallel" versions of a variety of image processing algorithms for our PIPE architecture. Based on our past experience with pipelined processors [41], we have already installed algorithms for spatial filtering, spatiotemporal filtering, and pyramid-based spatial processing [37]. Representative examples include edge preserving smoothing, generalized n-by-n convolution, normal optic-flow, thinning and morpological operators, and the pyramid representations of Burt, Mallat, and Singh and Ranganath [38].

4.3.2 Motion Perception Sensor Fusion

In work that is jointly supported by Philips Laboratories [39], we are implementing a multi-sensor fusion approach to the robust measurement of optic flow. Via confidence measures, we are integrating intensity correlation methods, which work best in structured scenes, and spatiotemporal energy methods, which are more suited for textured scenes [40, 42].

The spatio-temporal frequency approach is implemented on the PIPE using a pyramid image structure, called Pyramid of Oriented Edges; the POE is a logical extension of Burt's and of Mallat's pyramids, both of which we have also implemented. Using the POE, we have extracted coarse optic flow fields for a number of real images. We plan to extend the method by developing a hierarchical set of spatio-temporal frequency-tuned filters which will extract true optic flow from the POE data, and then integrate the results with our implementation of an intensity correlation-based model similar to that of Anandan.

4.3.3 Real-time Motion Tracking

The 60 Hz frame-rate image processing abilities of our PIPE enable it to generate visual tracking information fast enough to be coordinated with the motion control of a robotic arm. We have implemented pipelined algorithms to perform motion detection, thinning, and region-of-interest segmentation in order to track objects with a wrist-mounted camera in real-time [2]. Most of the processing is pyramid-based, and uses spatio-temporal filters. We have also implemented a motion-control process that concurrently calculates on the Masscomp host the predicted trajectory for the moving part, in order for the arm to intercept it for grasping.

5 Robotics and Tactile Sensing

We have recently acquired a Utah/MIT dextrous hand, for which we are developing tactile control algorithms. Having also recently acquired proprietary sensing "skin", we are also building its interface electronics and software. Through low-level sensor models and CAD/CAM object models, we continue to pursue the automatic generation of sensing constraints and strategies. We have begun to investigate the features, representations, and primitive operations necessary for the recognition of objects under multi-fingered and multi-jointed active haptic exploration.

5.1 System Development

5.1.1 Low-level Control

The Utah/MIT dextrous hand has provided us with a new set of tools to continue our study of intelligent touch and grasping. We are implementing Cartesian-based low level control algorithms for the hand, and extending them to a more hybrid scheme using both tendon force and tactile contacts.

5.1.2 Conformable Tactile Pads

We are currently implementing tactile sensors for each of the hand's fingers, using a proprietary piezoelectric polymeric material similar to that used on electronic drum pads. This responsive, conformable skin-like material can be deposited on arbitrary surfaces, and has extraordinarily good signal isolation and hysteresis characteristics. We are building a multiplexor for the sensor signals, in the hope of achieving real-time integration of the tactile sensor feedback into the low level hand control loop.

5.2 Multi-fingered Object Recognition

5.2.1 CAD/CAM and Sensor Models

Our experience with merging multiple sensor data sources [1, 3] has led us to examine the sensor modeling problem from the perspective of the automatic generation of viewpoint, geometric, and sensing constraints. We assume an assembly or an inspection domain, and our 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 sensors we model are, among others, monocular and binocular camera systems, laser range finders, and several types of active touch sensors.

5.2.2 Active Haptic Exploration

We have begun to analyze active multi-finger touch strategies to recognize objects haptically: that is, by only using external tactile sensors and internal force and position sensors. We are investigating the necessary data structures, procedural organizations, object models, and feature constraints that are the necessary prerequisites to active exploration; they may overlap with tools and methods in vision. We propose to demonstrate our haptic understanding of an object by establishing a secure enough grip to lift it.

6 References

1. Allen, P.K.. *Robotic Object Recognition Using Vision and Touch.* Kluwer Academic Publishers, 1987.

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

3. Allen, P.K. "Integrating Vision and Touch for Object Recognition Tasks". *International Journal of Robotics Research* (To appear 1988).

4. Boult, T.E. What is Regular in Regularization? Proceedings of the IEEE First International Conference on Computer Vision, June, 1987, pp. 457-462.

5. Boult, T.E., and Moerdler, M.L. An Experimental System for the Integration of Information from Stereo and Multiple Shape-from-Texture Algorithms. Proceedings of the SPIE Conference on Intelligent Robots and Computer Vision VI, November, 1987.

6. Boult, T.E., and Gross, A.D. Recovery of Superquadrics from Depth Information. Proceedings of the AAAI Workshop on Spatial Reasoning and Multi-Sensor Fusion, October, 1987, pp. 128-137.

7. Boult, T.E., and Gross, A.D. Recovery of Superquadrics from Depth Information. Proceedings of the SPIE Conference on Intelligent Robots and Computer Vision VI, November, 1987.

8. Boult, T.E. "Optimal Algorithms: Tools for Mathematical Modeling". *Journal of Complexity 3* (1987), 183-200.

9. Boult, T.E. Updating Distance Maps When Objects Move. Proceedings of the SPIE Conference on Mobile Robots II, November, 1987.

10. Boult, T.E. Regularization: Problems and Promises. Extended Abstracts of the 41st Annual SPSE Conference, May, 1988. (To appear).

11. Boult, T.E., and Chen, L.-H. Energy-Based Surface Segmentation. Department of Computer Science, Columbia University, February, 1988.

12. Boult, T.E., and Chen, L.-H. Analysis of Two New Stereo Algorithms Integrating Surface Reconstruction and Matching. Department of Computer Science, Columbia University, February, 1988.

13. Boult, T.E., and Gross, A.D. On the Recovery of Superellipsoids. Proceedings of the DARPA Image Understanding Workshop, April, 1988. (These proceedings).

14. Brown, L.G., and Shvaytser, H. Surface Orientation from Projective Foreshortening of Isotropic Texture Autocorrelation. Proceedings of the DARPA Image Understanding Workshop, April, 1988. (These proceedings).

15. 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. (To appear).

16. Chen, L.-H., and Boult, T.E. An Integrated Approach to Stereo Matching, Surface Reconstruction, and Depth Segmentation Using Consistent Smoothness Assumptions. Proceedings of the DARPA Image Understanding Workshop,

April, 1988. (These proceedings).

17. 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. (To appear).

18. Choi, D.J., and Kender, J.R. Solving the Depth Interpolation Problem on a Parallel Architecture. Proceedings of the IEEE Computer Society Workshop on Computer Architectures for Pattern Analysis and Machine Intelligence, October, 1987, pp. 107-114.

19. 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. (To appear).

20. Gross, A.D., and Rosenfeld, A. "Multiresolution Object Detection and Delineation". *Computer Vision, Graphics, and Image Processing 39* (1987), 102-115.

21. Hanvey, M. Environmental Representations for Mobile Robot Navigation. Department of Computer Science, Columbia University, January, 1988.

22. Hatzitheodorou, M., and Kender, J.R. An Optimal Algorithm for the Derivation of Shape from Shadows. Proceedings of the DARPA Image Understanding Workshop, April, 1988. (These proceedings).

23. 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. (To appear).

24. Ibrahim, H.A.H, Kender, J.R., and Shaw, D.E. "Low-Level Image Analysis Tasks on Fine-Grained Tree-Structured SIMD Machines". *Journal of Parallel and Distributed Computing 4* (1987), 546-574.

25. Ibrahim, H.A.H. Pyramid Algorithms Implementation on the Connection Machine. Proceedings of the DARPA Image Understanding Workshop, April, 1988. (These proceedings).

26. Kender, J.R., and Smith, E.M. Shape from Darkness: Deriving Surface Information from Dynamic Shadows. Proceedings of the IEEE First International Conference on Computer Vision, June, 1987, pp. 539-546.

27. Kender, J.R., and Freudenstein, D.G. What is a Degenerate View? Proceedings of the Tenth International Joint Conference on Artificial Intelligence, August, 1987, pp. 801-804.

28. Kender, J.R., and Leff, A. On the Computational Complexity of Linear Navigation. Proceedings of the DARPA Image Understanding Workshop, April, 1988. (These proceedings).

29. Kender, J.R., and Smith, E.M. A Method and an Apparatus for Determining Surface Shape Utilizing Object Self-Shadowing. Patent Application 058,914. July, 1987.

30. Moerdler, M.L., and Kender, J.R. An Integrated System that Unifies Multiple Shape-from-Texture Algorithms. Proceedings of the Sixth National Conference on Artificial Intelligence, July, 1987, pp. 723-727.

31. Moerdler, M.L., and Kender, J.R. An Approach to the Fusion of Multiple Shape-from-Texture Algorithms. Proceedings of the AAAI Workshop on Spatial Reasoning and Multi-Sensor Fusion, October, 1987, pp. 272-281.

32. Moerdler, M.L., and Boult, T.E. The Integration of Information from Stereo and Multiple Shape-from-Texture Cues. Proceedings of the DARPA Image Understanding Workshop, April, 1988. (These proceedings).

33. Moerdler, M.L., and Boult, T.E. The Integration of Information from Stereo and Multiple Shape-from-Texture. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June, 1988. (To appear).

34. Roberts, K.S., and Ganapathy, S.K. Stereo Triangulation Techniques. Proceedings of the IEEE Computer Society Workshop on Computer Vision, November, 1987, pp. 336-338.

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

36. 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. (To appear).

37. Singh, A. Image Processing on PIPE. Robotics and Flexible Automation Department, Philips Laboratories, June, 1987.

38. Singh, A., and Allen, P.K. Constructing Multi-Resolution Image Representations in Real Time. Department of Computer Science, Columbia University, July, 1987.

39. Singh, A., and Ranganath, S. A Hierarchical Model for Directional Selectivity in space and Measurement of Optic-Flow. Robotics and Flexible Automation Department, Philips Laboratories, August, 1987.

40. Singh, A. Measurement of Visual Motion. Department of Computer Science, Columbia University, October, 1987.

41. Singh, A., and Lala, P.K. "A Multilayer Cellular Architecture for a Highly Parallel VLSI Supercomputer". *IEEE Transactions on Computers* (To appear 1988).

42. Singh, A., and Allen, P.K. A Real-Time Hierarchical Model for Optic Flow Determination via Spatiotemporal Frequency Channels. Proceedings of the DARPA Image Understanding Workshop, April, 1988. (These proceedings).

43. Wolberg, G. "A Syntactic Omni-font Character Recognition System". *International Journal of Pattern Recognition and Artificial Intelligence 1* (1987), 303-322.

44. Wolberg, G. Image Warping Among Arbitrary Planar Shapes. Proceedings of the Computer Graphics International Conference, May, 1988. (To appear).

45. Wolff, L.B. A General Formalization of Stereo Vision. Proceedings of the SPIE Conference on Intelligent Robots and Computer Vision VI, November, 1987.

46. Wolff, L.B. Spectral and Polarization Stereo Methods Using a Single Light Source. Proceedings of the IEEE First International Conference on Computer Vision, June, 1987, pp.

708-715.

47. Wolff, L.B. Surface Orientation from Polarization Images. Proceedings of the SPIE Conference on Optics, Illumination, and Image Sensing for Machine Vision II, November, 1987.

48. Wolff, L.B. Shape from Polarization Images. Proceedings of the IEEE Computer Society Workshop on Computer Vision, November, 1987, pp. 79-87.

49. Wolff, L.B. Towards a More General Conception of Stereo Vision. Proceedings of the AAAI Spring Symposium Series, March, 1988.

50. Wolff, L.B. An Introduction to Generalized Stereo Techniques. Proceedings of the DARPA Image Understanding Workshop, April, 1988. (These proceedings).

51. Wolff, L.B. Ray Tracing Using Polarization Parameters. Department of Computer Science, Columbia University, January, 1988.