AD-A278 540

MENTATION PAGE

Form Approved
OMB No. 0704-0188

(2)

...estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, ...ng and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this ...ng this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson ...to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

REPORT DATE
1/11/94

3. REPORT TYPE AND DATES COVERED
Final 15Sep90 - 31MAR94

4. TITLE AND SUBTITLE
RESEARCH IN IMAGE UNDERSTANDING (u)

5. FUNDING NUMBERS
F 49620-90-C-0078
61102F
7515/00

6. AUTHOR(S)
Ramakant Nevatia

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Southern California
Institute for Robotics & Intelligent Systems
Computer Science Dept., PHE 204
Los Angeles, CA 90089-0273

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR- 94 0248
IRIS Report No.
94-320

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
ARPA
Attn: Program Manager
3701 No. Fairfax Drive
Arlington, VA 22203-1714

USAF, AFSC
AFOSR
Building 410
Bolling AFB DC 20332-6448

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

DTIC
ELECTE
APR 2 5 1994
S G D

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE: Distribution Unlimited

12b. DISTRIBUTION CODE
UL

13. ABSTRACT *(Maximum 200 words)*

We undertook a broad program for research into image understanding techniques suited for a variety of applications. We divided our tasks into three major categories. However, we wish to emphasize that the different tasks are highly interelated and share many common techniques. The major task areas over the course of this contract were three-dimensional vision including descriptions from range data, shape inference from images, and object recognition; motion analysis and parallel processing. This report discusses the status of the various individual research projects funded by this contract.

14. SUBJECT TERMS
Computer vision; image analysis, three-dimensional descriptions, motion estimation; mobile robot

15. NUMBER OF PAGES
20

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL (SAR)

NSN 7540-01-280-5500

DTIC QUALITY INSPECTED 3

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

Research in Image Understanding

Contract F49620-90-C-0078

FINAL TECHNICAL REPORT

September 1, 1990 - September 30, 1993

Institute for Robotics and Intelligent Systems
Powell Hall Room 204
University of Southern California
Los Angeles, CA 90089-0273

Research in Image Understanding

Contract F49620-90-C-0078
Final Technical Report
September 1, 1990 to September 30, 1993
DARPA Order A.O. 7515/00
Program Code 0E20
Contractor: University of Southern California
Start Date: 9/01/90
Expiration Date: 03/31/93
Contract Dollars: $1,773,472
Principal Investigator: Ram Nevatia
(213) 740-6427
Program Manager: Abraham Waksman
(202) 767-5025

R. Nevatia and K. Price (Editors)
Institute for Robotics and Intelligent Systems
School of Engineering
University of Southern California
Los Angeles, CA 90089-0273

November, 1993

Sponsored by Defense Advanced Research Projects Agency DARPA Order No. 7515/00 Monitored by AFOSR Under Contract No. F49620-90-C-0078

1 INTRODUCTION

This report summarizes research progress over the term of the contract. We will give a brief overview of the work in this introduction with more detail provided in later chapters. Longer discussions of many of these projects has appeared in the Image Understanding Workshop Proceedings for January 1992 and April 1993. We also provide references to these and other published references to this work. Additionally, many of the papers have appeared in the Annual Technical reports for 1991, 1992 or 1993.

We undertook a broad program for research into image understanding techniques suited for a variety of applications. We divided our tasks into three major categories. However, we wish to emphasize that the different tasks are highly interrelated and share many common techniques. The major task areas over the course of this contract were three-dimensional vision including descriptions from range data, shape inference from images, and object recognition; motion analysis and parallel processing. The introduction will briefly outline the different task areas from the proposal with discussion of the status of work on that topic. These tasks are discussed in more detail in the following sections.

1.1 Three Dimensional Vision

The ability to describe and recognize 3-D objects is needed for many tasks including those of manufacturing robotics as well as for outdoor object recognition. For the project our goals included:

- Develop formal methods to represent 3-D objects, in their entirety or their visible faces only, in terms of volumetric or surface descriptors for polyhedra and free form sculptured objects.
- Develop techniques to robustly compute these descriptors in real images in the presence of noise, shadows, surface markings and occlusion.
- Develop techniques that process input from different sources including range, stereo and monocular intensity images.
- Develop techniques that use the description of 3-D objects for robust recognition and pose identification, taking advantage of the richness of our high-level symbolic descriptions to perform indexing into the database of stored objects.
- Develop a system to automatically acquire the descriptions of a single 3-D object model from a series of views.

These goals were addressed in several research projects:

- Developed a system to generate three-dimensional descriptions from range data using deformable models to describe arbitrary shapes, see Section 2.1.2 and [28,32].
- Developed programs to generate 3-D descriptions from single gray scale images based on analysis of contours, see Section 2.4 and [66, 67]. This work involves theoretical analysis of the appearance of contours and implementations of these theories.
- Developed a system for generating segmented 3-D descriptions from gray scale images using symmetries, see Section 2.1.3 and [54]
- Developed 3-D descriptions from stereo views, especially applied to buildings [7,8].
- Developed general perceptual grouping techniques that have been as an important part of this work [15]
- Used three-dimensional and two-dimensional descriptions of objects for recognition is several task domains, see Section 2.2 and [56]
- Developed two complementary systems that combine range images from several different directions to produce a complete model of the object. The first combines the range images to generate a complete description [1,3]. The second approach generates descriptions for each view and combines these into a boundary based description of the objects [38]. See Section 2.1.1 for more information.

1.2 Motion Analysis

This includes detection and description of moving objects and inference of the three-dimensional structure of the environment. We propose to develop the fundamental parts needed in a general motion understanding system and to demonstrate their viability on some specific scenarios. Our goals in the analysis of sequences of images included:

- Develop techniques for estimating 2-D (image plane) motion and object segmentation in the presence of occlusion in closely spaced image sequences (i.e. spatio-temporal data) and for inferring the object structure from the data.
- Develop techniques for matching contours and other features that apply to broad classes of scenes and motions. This will include the use of feedback from higher levels of processing such as motion estimation and 3-D inference.
- Develop a system for effective 3-D inference from motion sequences combining the computation of depth and structure from motion with other general techniques and for merging local 3-D scene descriptions into global descriptions.
- Develop an integrated motion analysis system and test on increasingly difficult scenarios. The system will use existing analysis modules, but will include interaction between systems and feedback of results to the feature extraction process.

Through several projects we addresses these goals:

- Developed a slice-based analysis technique that computes dense optical flow estimates for arbitrary observer motion in closely spaced data, see Section 3.1 and [43].
- Developed a program for contour matching in image sequences [13] and a second system for matching both region and corner features within a larger motion analysis system, see Section 3.3 and [24].
- Developed a three-dimensional motion estimation system that uses an arbitrary sequence (at least three frames) of matched points. This system allows for constant motion and a limited case of accelerations along with handling occlusions and missing matches. See Section 3.2 and [11].
- Developed an integrated system that combined automatic feature extraction and matching with the 3-D motion estimation system to produce 3-D descriptions of detected objects. The combination included feedback of motion and structure estimates to the matching to eliminate poor matches and find missing features. This work is discussed more fully in Section 3.3 and [26].
- Developed a trinocular (three camera) vision system for use on a mobile platform vision for guidance and obstacle avoidance. The three cameras simplify the stereo computations and produce rapid reliable estimates of structures in indoor environments, see Section 3.4 and [23].

1.3 Parallel Processing

This includes work in mapping existing algorithms to available parallel implementations, and the study of how parallel architectures can be designed to aid in the development of parallel algorithms. The development of parallel techniques will support efficient implementations of systems developed in the other tasks. Our initial goals were:

- Develop a mapping for image analysis algorithms onto massively parallel architectures.
- Develop techniques for describing the communication necessary in parallel implementations of symbolic image analysis algorithms.

We addressed these in two projects in mapping computer vision algorithms to parallel architectures:

- We analyzed the communication requirements for implementing computer vision algorithms in a variety of existing parallel architectures. This is discussed in Section 4.1 and [22, 50].
- Developed a general analysis technique to study different massively parallel architectures and their application to a variety of computer vision algorithms. This is described in Section 4.5 and [49,51].

2 THREE-DIMENSIONAL VISION

2.1 Description of 3-D Objects

2.1.1 Integration from Multiple Views

We have developed systems for building models from unregistered multiple range images [39, 2]. The latter system integrates views at the triangulated surface level rather than at the pixel level. A triangulated surface model can represent a variety of solid objects, and theoretically to any kind of resolution. They are not ideal representations for high level vision tasks, such as recognition, because, first, the representation is still low level, second, it is sensitive to many parameters, and therefore unstable. However, we think it is a good intermediate representation for integration and for building high level description through surface interpolation from triangulation.

2.1.2 Deformable Models

A second project in range analysis involves the use of deformable surfaces to generate a 3-D approximation of range data. This work, performed by C. Liao and G. Medioni, builds on our earlier work in "B-Snakes." The user provides an initial simple surface, such as a cube, which is subject to internal forces (describing implicit continuity properties such as tension and bending) and external forces which attract it toward the data points. The problem is cast in terms of energy minimization. We solve this non-convex optimization problem by using the well known Powell algorithm which guarantees convergence to a (possibly local) extremum and does not require gradient information. The variables are the positions of the control points. The number of control points is adaptively controlled. This methodology leads to a reasonable complexity and good numerical stability. We also provide a novel solution to the problem of subdividing a patch when the fit is bad. We show results on real range images to illustrate the applicability of our approach. The advantages of this approach are that it provides a compact representation of the approximated data, and lends itself to applications such as non-rigid motion tracking and object recognition. Currently, our algorithm gives only a C^0 continuous analytical description of the data, but due to the flexibility of our adaptive approach it should be upgraded to C^1 or C^2 easily. This work is discussed in more detail in [28].

2.1.3 Segmented Volumetric 3-D Descriptions

We address the problem of recovering segmented hierarchical volumetric descriptions of three dimensional shapes. In an earlier work [52,54], we have suggested a method (using SLS) for obtaining hierarchical axial descriptions of planar shapes, together with a decomposition of the shapes into their parts. Unfortunately, it is not straightforward to extend these methods to handle three dimensional shapes. This is because in the three dimensional space the SAT and SLS axes are, in general, not curves, but surfaces, leading to unnatural descriptions [79].

In this current work, performed by H. Rom and G. Medioni, we restrict ourselves to three types of parts: Convex blobs (or Ovoids, borrowing the terminology from Koenderink [75]), Straight Homogeneous Generalized Cylinders (SHGCs [82]), and Planar Right Constant GCs (PRCGCs [63], planar axis and constant cross section). These components exhaust many of the man-made objects encountered on a normal basis. We suggest the use of properties of the parabolic curves (zero crossings of the Gaussian curvature) for recovering the cross sections and axes of the different parts. We advocate the use of the parabolic curves over the often used occluding contours, which are unstable in range data. We will assume that the shapes are C^2 continuous (i.e. the curvature is defined everywhere). We do not want to assume, as several authors do, that the parts are cut along a cross section or that a cross section is visible. Furthermore, we will not assume the existence of any discontinuity edges between parts. We believe that the case of parts joined discontinuously is the limiting case of the more general continuous case which we address.

Given the 3-D surface data, either from a CAD model, or from registered range images [2], or from a single range image, we first recover the parabolic curves on the surface. This requires the evaluation of the sign of the Gaussian curvature of the surface patches. It has been shown that this process is stable and reliable [70,80,9]. The parabolic curves could be either on the surface of the individual parts, or on the border of the “glue” between parts. Note, that due to the transversality principle [72], there is almost always an anticlastic (negative Gaussian curvature) region between convex parts when they are joined. The parabolic curves on the parts we consider could be either meridians or cross sections of the SHGC and PRCGC parts (this has been shown for SHGCs [81] and we have proven it for PRCGCs). Using simple tests we can hypothesize (or in many cases determine) the role of each parabolic curve. We can therefore segment the object into parts, and based on the properties of the specific parts, we can recover the axis of the parts from the meridians and cross sections.

One problem which remains is that some parts cannot be found until some other parts are removed. As in [52] and [54], we take an hierarchical strategy, in which, at each step, well defined parts are described and removed. Once these parts are removed, the next level parts can now be described. This process is efficient and produces a decomposition of the shape into its intuitive parts with a stable axial description of these parts.

2.2 Object Recognition

The more interesting problem is to recognize 3D objects from grey level images. The previous methodology becomes very inefficient, as the number of generated hypotheses increases drastically. We propose instead to generate high level groupings and to use these as matching primitives. The groupings we are using are based on parallel and skew symmetry, U-shapes and closures. Furthermore, we show how to compute these groupings efficiently from segments, and how we keep the number of groupings small. We have obtained encouraging initial results on real images [7].

As an application of our matching methodology, we study the "drop-off" problem, in which an observer is given a topographic map, and is dropped off at an unknown location. We select as matching feature the panoramic horizon curve (corresponding to the sky-ground boundary from a given viewpoint). The polygonal approximation of this curve is compared with precomputed ones using our hash based scheme [56]. We have obtained accurate results from real data.

We have defined a methodology based on efficient coding and hash tables to recognize 3D objects given 3D data, even when the number of models is large [57]. We have performed a detailed complexity analysis of the method, which results in $O(n) \leq O(recognition) \leq O(nm^3)$, where n is the number of matching primitives and m is the number of models in the database. The worst case occurs when the models hypothesized to be in the scene are very similar.

2.3 Perceptual Grouping

Most high level vision algorithms, such as shape from contour [60] or line drawing interpretation require perfect data as input, but it is impossible to generate such features with low level algorithms such as edge detectors. Here, we try to bridge this gap by transforming an edge image into a saliency map. This approach uses a non-iterative method based on a field associated with each edge. This field encodes the notions of simplicity, curvature constancy and co-curvilinearity. A detailed report on this effort is given in [15].

2.4 Shape Analysis from Monocular Images.

In this project, we are developing techniques for inferring 3-d shape descriptions given only object contours. We have developed a theory that can infer the shape of a class of objects, namely zero-Gaussian curvature surfaces, straight homogeneous generalized cylinders and planar, right, constant cross-section generalized cylinders. One of the recent advances here is extension of our techniques to infer shape of objects made of multiple curved surfaces. We have also started to develop techniques to make our theory work with real images where contours are likely to be fragmented and distracting contours such as markings and shadows present. We report on these efforts in [66, 67].

We have continued our effort in understanding how to infer shape from monocular images using contours. First we developed a theory of invariances of projected contours and how they can be used to infer 3-D shapes of a certain classes of surfaces [61, 64, 65]. In recent work, we have developed a system for generating volumetric 3-D shape descriptions from real images containing Straight, Homogeneous Generalized Cylinders (SHGCs). The image may contain multiple, occluding objects and the objects may have surface markings. In working with real images, we must deal with problems of fragmented boundaries and many additional boundaries due to markings, shadows, highlights and noise. We use the expected properties of the desired contours to separate the two sets of properties and to complete the broken boundaries. Details of this process are given in a paper in the recent IU Workshop [66].

In continuation of this work, we are also studying the class of curved generalized cylinders with circular but changing cross-sections. In this case, we are unable to find invariants for the visible boundaries. However, we are able to find good *quasi-invariants* that show that commonly used ribbon descriptions are in fact, stable for such objects. Our future work will focus on compound objects that combine a number of primitives that we have analyzed in the past.

3 MOTION ANALYSIS

3.1 Spatio-Temporal Analysis

The goal of our work in spatio-temporal analysis is to generate a dense optic flow map from a motion sequence. Because of the sparseness of 0D features (e.g. corners) or 1D features (e.g. curves), we feel that 2D features (e.g. regions) are more likely to produce dense motion estimates. Early work in spatio-temporal analysis includes that of [71]. Our work began with [41,42], with the extraction of paths in slices taken in the temporal direction of the spatio-temporal data volume (i.e. paths of an object point through time and space). This produces an image velocity estimate only along object contours.

In order to generate a dense displacement field, more analysis of the slice data is needed. Strips that correspond to trapezoidal regions found in the slices through the temporal dimension of the image volume are constructed for selected orientations throughout the image. These extracted strips provide estimates of the velocity component along the slice orientation. The velocity estimates of different slice orientations are combined to compute the velocity constraint for each pixel. A voting scheme is used to extract the position of the Focus of Expansion, which can then be used to compute the real velocity of the pixels.

This process is very expensive (requiring hours on serial machines), but most of the computation is easily performed on the SIMD architecture of the Connection Machine. This algorithm was transferred to a CM-2 with very good results for computational speed-up. Much of this work was reported on in previous years and has now been completed with more detail in [43].

3.2 Motion Estimation

We have continued our exploration of techniques for computing structure from motion using feature matches through multiple frames. The use of multiple (as opposed to two) frames is desirable for several reasons:

- to increase the robustness of the solution,
- to allow recovery of structure/motion with fewer features being tracked, and
- to allow estimation of "higher order derivatives" of the motion.

We have completed development and implementation of an algorithm for the shape from motion problem given point feature correspondences and perspective projection. This solution works for a class of motions called chronogeneous motion, which includes uniform acceleration and constant angular velocity rotation and translation as special cases. The solution is by an iterative algorithm that recovers the three-dimensional motion of the feature points and the three-dimensional location of each feature in each frame. An additional closed form algorithm that recovers motion and structure for uniform acceleration is used to generate initial guesses for the iterative procedure [10].

These algorithms are discussed further in [12] with additional results, or in the thesis [11]. The results show that this algorithm performs well in recovering structure and motion parameters from feature point correspondences. We are using this motion estimation technique in our other motion work [24].

3.3 Integrated Motion System

Accurate motion estimation in feature-based analysis of an image sequence requires consistent feature extraction and reliable matching. Without *a priori* information, inconsistent feature extraction and erroneous matching are hard to detect and are often closely related. To address these problems, we developed an integrated feature-based (including both image regions and corner features) system that uses errorful data for motion estimation and overcomes these errors with feedback that improves both matching and feature extraction. The system acquires the initial estimates using a batch mode analysis (as in [11]) and continues the processing using an incremental analysis of the input sequence. Thus we use three dimensional motion estimation as an aid in generating the data necessary for the motion estimation system itself. The program refines the initial noisy correspondence data by removing those parts that do not fit the estimated 3-D motion parameters. The motion parameters are in turn refined using the improved correspondence data. This process is iterated until a consistent feature set is obtained (usually two to three times). For improved feature extraction, properties obtained from the corresponding object in other frames guide the segmentation process to improve the extracted region or add missing regions to the sequence of tracked features. By using the regions which underlie the corners, we produce a rough reconstruction of surfaces in the environment from the sparse depth information at the corners.

This system used several existing programs for feature extraction and matching and for estimating three dimensional motion and structure from a set of matching points. These systems have limitations and produce errors such as missed or extra features, missed or incorrect matches. By incorporating these different programs into a single system with feedback, we have substantially reduced the impact of the errors and improved the final results of the analysis. In the papers and thesis on this work, we present the results on standard real image sequences, which are a subset of the data the program has been tested on. More details are presented in [24,26,27]

3.4 Mobile Platform

We have continued with our robot project using trinocular imagery for guidance. We are investigating robot navigation for situations where only *generic* maps are available, with one of the tasks being the generation of more complete maps. The visual navigation uses three views to improve the performance of the stereo system, both in speed and accuracy of the matching. Rather than producing a complete depth map we are concerned only with producing a "squeezed 3-D map" that shows corridor walls and obstacles in the hallway. This work is discussed in more detail in [23]

4 PARALLEL PROCESSING

4.1 Algorithm Implementation on Existing Machines

We have studied parallel implementations of several high-level algorithms, such as relaxation labelling and graph matching. Our recent work has looked at the problem of geometric hashing, which is used for a variety of matching problems [58]. In earlier parallel implementations the number of processors was independent of the size of the scene but depended on the size of the model database. In this work we have designed new parallel algorithms for both the MasPar and Connection Machine architectures which improve on the number of processors and improve the overall performance. Details of this work are given in [22]. A summary of our recent work is outlined below.

4.1.1 Stereo and Image Matching

Stereo matching is one of the well known methods for extraction of depth information. Depth recovery is a crucial problem in image understanding with applications in robotics and navigation. For stereo matching, we have proposed $O(Nn^3/P)$ time algorithm on a P processor fixed size linear array, where N is the number of line segments in one image, n is the number of line segments in a window determined by the object size, and $P \leq n$ [21]. This algorithm is a parallel implementation of the stereo matching algorithm proposed by Medioni and Nevatia in [31].

Discrete relaxation techniques have been widely used in computer vision and artificial intelligence. For the image matching problem, discrete relaxation technique outlined in [30] leads to a sequential execution time of $O(n^3m^3)$ for labelling n objects with m labels. In [29] we have proposed a faster sequential algorithm for image matching which runs in $O(n^2m^2)$ time, where n is the number of line segments in the image and m is the number of line segments in the model. Also, a partitioned parallel implementation has been developed by using the proposed sequential algorithm. $O((nm/P+P)nm)$ time performance is achieved on a P processor fixed size linear array, where $P \leq nm$.

4.1.2 Sorting on Reconfigurable Mesh

The Reconfigurable Mesh forms the CAAPP level of Image Understanding Architecture (IUA) [83]. An optimal sorting algorithm on the Reconfigurable Mesh is derived in [20]. The algorithm sorts n numbers in constant time using $n \times n$ processors. The best known previous result uses $O(n \times n \log^2 n)$ processors. Our algorithm satisfies the AT^2 lower bound of $\Omega(n^2)$ for sorting n numbers in the word model of VLSI. Modification to the algorithm for area-time trade-off is shown, to achieve the AT^2 lower bound over $1 \leq T \leq \sqrt{n}$. Previously, the lower bound was achieved over $\log n \leq T \leq \sqrt{n}$. Notice that, using sort as a basic procedure number of low- and intermediate-level Image Understanding problems can be solved on the IUA.

4.1.3 Graph Algorithms

Many of the intermediate-level computer vision tasks can be posed as graph problems. Particularly, digitized picture graphs (DPGs) of two and three dimensions are of primary importance due to their natural correspondence with black/white images. We introduce a notion of *partitionability* of graphs and show that DPGs (of any fixed dimension) are partitionable [45]. This partitionability property helps in constructing efficient parallel algorithms for many problems on digitized picture graphs. We show that our techniques can be efficiently simulated on a $P \times P$, fixed-size mesh-connected computer, $1 \leq P \leq n$. Unlike other approaches [78], our algorithms, because of the partitionability idea, easily extend to problems in higher dimensions.

4.1.4 VLSI Architectures for Image Transforms and Vector Quantization

We have studied VLSI architectures for various image transforms and vector quantization techniques. Two linear array architectures have been proposed for computing the arithmetic Fourier transform and image compression using vector quantization [35,36]. These architectures have modular PEs and can support real-time processing. The designs can operate with less number of PEs than the input size. The proposed designs require fixed I/O bandwidth with the host.

4.2 Parallelization of Symbolic Techniques in Vision

There is relatively little work done in parallelizing high level vision algorithms. Such algorithms are usually symbolic in nature and the processing is not entirely local. We believe that when dealing with such complex algorithms, the parallel implementation must be concerned with the following four characteristics: algorithm speedup, processor efficiency, system complexity, and programmer burden.

Most research in parallel processing has been concerned solely with the first two characteristics. We have been pursuing an alternative that achieves a better balance between the desired characteristics. In our approach we classify algorithms, in terms of operations, data dependencies, data movements, and algorithm characteristics, and then specify a parallel processor architecture that is well suited to those characteristics [51].

We have applied this methodology to a number of mid and high level vision algorithms. Our first experience was with an algorithm for image matching via relaxation labelling with symbolic objects and geometric constraints[30]. Our analysis indicated that the use of an MIMD architecture that comprises powerful processing elements programmed with the loosely synchronous protocol. A suitable interconnect topology is one of logarithmic diameter. Two implementations were developed, one using binary tree and the other using hypercube connections. This scheme exploits the coarse grain parallelism within the algorithm. Further analysis shows that equipping each PE with a tightly coupled vector processor would exploit the fine grain parallelism within the algorithm. This architecture achieves high degrees of speedup and efficiency while using software that is nearly identical to that of the serial implementation, thus system complexity and programmer burden are minimized. Details of this study were previously reported in [49].

In more recent work, we have studied an algorithm for object recognition that uses graph matching. Our analysis again indicates that an MIMD architecture with powerful processing elements is suited to this problem. The PEs are connected by a hypercube topology. Again, we achieve significant algorithm speedup and processor efficiency while using software that is nearly identical to that of the serial implementation.

Lastly, we have studied the mid-level operations of linear feature extraction and perceptual organization. These operations may appear to be simple and repetitive, and thus well suited to SIMD implementations. However, this is not the case. Typically, parallel implementations only focus on the study of a specific algorithm. They assume that the input is given in the desired form and the output is produced in some form. In a system (or sub-system) that comprises of a number of processing steps, conversion of the output of one stage to another itself can be a major step, possibly requiring serial implementation. A simple example is that of linear feature extraction where finding edges and then their neighbors that would form curves is an iconic process that is easily implemented on a SIMD machine. However, this is different from actually producing a *list* of curves, each curve given by a list of points forming it, in order, and possibly a linear approximation to it as well. This is the structure needed for subsequent use of the linear feature processing. Our proposed implementation is described in detail in [50].

In conclusion, our research has shown that the complex operations and data movements required by mid and high level vision can be performed efficiently if care is taken in specifying the parallel processor architecture. Implementations that achieve high degrees of algorithm speedup and processor efficiency can be attained without sacrificing system complexity and programmer burden. Such architectures can be realized utilizing heterogeneous designs or via reconfigurable architectures given an efficient reconfiguration procedure. We have certainly not studied all the algorithms used in vision, but believe that our choice of selected algorithms covers enough of a span to indicate that it is fruitful to further pursue this approach. We also believe that the next step is to investigate the parallel implementation of complete, heterogeneous vision systems that comprise low, mid, and high-level algorithms.

5

References

5.1 Publications from Contract

[1] Y. Chen and G. Medioni, "Object Modeling by Registration of Multiple Range Images," In *Proceedings of the IEEE Conference on Robotics and Automation*, volume 3, pages 2724–2729, April 1991.

[2] Y. Chen and G. Medioni, "Object Modelling by Registration of Multiple Range Images," *International Journal of Image and Vision Computing*, 10(3):145–155, April 1992.

[3] Y. Chen and G. Medioni, "Surface Level Integration of Multiple Range Images," in *Proc. DARPA Image Understanding Workshop*, Washington, DC, April 1993

[4] Y. Chen and G. Medioni. "Surface Level Integration of Multiple Range Images," In *Proceedings of the Workshop on Computer Vision for Space Applications*, Antibes, France, September 1993.

[5] C.-K. R. Chung and R. Nevatia, "Use of Monocular Groupings and Occlusion Analysis in a Hierarchical Stereo System," In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 50–56, Maui, Hawaii, June 1991.

[6] C.-K. R. Chung and R. Nevatia, "Recovering LSHGCs and SHGCs from Stereo," In *Proceedings of the DARPA Image Understanding Workshop*, pages 401–407, San Diego, CA, January 1992.

[7] C.-K. R. Chung and R. Nevatia, "Recovering Building Structures from Stereo," In *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pages 64-73, Palm Springs, CA, December 1992.

[8] C.-K. R. Chung, *Deriving 3-D Shape Descriptions from Stereo Using Hierarchical Features*, Ph.D. Thesis, University of Southern California, November 1992.

[9] T. J. Fan, G. Medioni, and R. Nevatia. "Recognizing 3-D Objects Using Surface Descriptions,"*IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1140–1157, 1989.

[10] W. Franzen, "Structure and Motion from Uniform 3D Acceleration," In *Proceedings of the Workshop on Visual Motion*, pages 14–20. IEEE Computer Society, Oct 7–9, 1991.

[11] W. Franzen, *Structure from Chronogeneous Motion*, Ph.D. thesis, University of Southern California, May 1991, IRIS Technical Report 266.

[12] W. O. Franzen, "Structure from Chronogeneous Motion: A Summary," In *Proceedings of the DARPA Image Understanding Workshop*, San Diego, CA, January 1992.

[13] S. L. Gazit and G. Medioni, "Multi-Scale Contour Matching in a Motion Sequence," in *Proc. DARPA Image Understanding Workshop*, Palo Alto, CA, May, 1989, pp. 934-943.

[14] G. Guy and G. Medioni, "Perceptual Grouping Using Global Saliency Enhancing Operators," In *Proceedings of the International Conference on Pattern Recognition*, pages 99–104, The Hague, Holland, August 1992.

[15] G. Guy and G. Medioni, "Inferring Global Perceptual Contours from Local Features," in *Proceedings of the DARPA Image Understanding Workshop*, Washington, DC, April 1993.

[16] A. Huertas and R. Nevatia, "Detecting Buildings in Aerial Images," *Computer Vision, Graphics, and Image Processing*, 41(2):131–152, February 1988.

[17] A. Huertas, W. Cole, and R. Nevatia, "Using Generic Knowledge in Analysis of Aerial Scenes: A Case Study," In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1642–1648, Detroit, MI, August 1989.

[18] A. Huertas, W. Cole, and R. Nevatia, "Detecting Runways in Complex Airport Scenes," *Computer Vision, Graphics, and Image Processing*, 51(2):107–145, August 1990.

[19] A. Huertas, C. Lin and R. Nevatia, "Detection of Buildings from Monocular Views of Aerial Scenes Using Perceptual Grouping and Shadows," in *Proceedings of the DARPA Image Understanding Workshop*, Washington, DC, April 1993.

[20] J. Jang and V. K. Prasanna, "An Optimal Sorting Algorithm on Reconfigurable Mesh," Technical report, University of Southern California, Los Angeles, 1991.

[21] A. Khokhar, W. Lin, and V. K. Prasanna, "Stereo and Image Matching on Fixed Size Mesh Based Arrays," In *Computer Architectures for Machine Perception Workshop*, 1991.

[22] A. Khokhar and V. Prasanna, "Scalable Data Parallel Geometric hashing: Experiments on MasPar MP-1 and on Connection Machine CM-5," In *Proceedings of the DARPA Image Understanding Workshop*, Washington, DC, April 1993.

[23] D. Kim and R. Nevatia, "Indoor Navigation without a Specific Map," in *Proceedings Workshop on Intelligent Systems*, Pittsburgh, PA, 1993

[24] Y. C. Kim and K. Price, "Improved Correspondence in Multiple Frame Motion Analysis," In *Proceedings of the DARPA Image Understanding Workshop*, San Diego, CA, January 1992.

[25] Y. Kim and K. Price, "Refinement of Noisy Correspondence Using Feedback from 3-D Motion," *Proceedings of the Conference on Computer Vision and Pattern Recognition*, June 1992, Champaigne, IL. pp. 836-838.

[26] Y. C. Kim and K. Price, "A Feature-Based Monocular Motion Analysis System guided by Feedback Information," In *Proceedings of the DARPA Image Understanding Workshop*, Washington, DC, April 1993.

[27] Y. C. Kim, *An Integrated Feature-Based Monocular Motion Analysis 'System Using Feedback*, Ph. D. Thesis, University of Southern California, August 1993.

[28] C. Liao and G. Medioni, "Adaptive Surface Approximation of a Cloud of 3D Points," USC report, submitted for publication.

[29] W. Lin and V. K. Prasanna, "Parallel Algorithms and Architectures for Discrete Relaxation Technique," In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 1991.

[30] G. Medioni and R. Nevatia, "Matching Images Using Linear Features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):675–685, November 1984.

[31] G. Medioni and R. Nevatia, "Segment-Based Stereo Matching," *Computer Graphics and Image Processing*, 31(1):2–18, July 1985.

[32] S. Menet, P. Saint-Marc, and G. Medioni, "B-Snakes: Implementation and Application to Stereo," In *Proceedings of the DARPA Image Understanding Workshop*, pages 720–726, Pittsburgh, Pennsylvania, September 1990.

[33] R. Mohan and R. Nevatia, "Segmentation and Description Based on Perceptual Organization," In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 333–341, San Diego, California, June 1989.

[34] R. Mohan and R. Nevatia, "Using Perceptual Organization to Extract 3-D Structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1121–1139, November 1989.

[35] H. Park and V. K. Prasanna, "VLSI Architectures for Computing the Arithmetic Fourier Transform," In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1991.

[36] H. Park and V. K. Prasanna, "VLSI Architectures for Real Time Vector Quantization," In *International Symposium on Circuits and Systems*, 1991.

[37] B. Parvin and G. Medioni, "A Dynamic System for Object Description and Correspondence," In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 393–399, Maui, Hawaii, June 1991.

[38] B. Parvin and G. Medioni, "A Layered Network for the Correspondence of 3D Objects," In *Proceedings of the IEEE Conference on Robotics and Automation*, pages 1808–1813, April 1991.

[39] B. Parvin and G. Medioni, "B-rep from Unregistered Multiple Range Images," In *Proceedings of the IEEE Conference on Robotics and Automation*, Nice, France, May 1992.

[40] B. A. Parvin, *A Dynamic System for Object Description and Correspondence*, Ph.D. thesis, University of Southern California, July 1991. "IRIS Technical Report 286.

[41] S.-L. Peng and G. Medioni, "Spatio-Temporal Analysis of Velocity Estimation of Contours in an Image Sequence with Occlusion," In *Proceedings of the International Conference on Pattern Recognition*, pages 236–241, Rome, Italy, November 1988.

[42] S.-L. Peng and G. Medioni, "Interpretation of Image Sequences by Spatio-Temporal Analysis," In *Proceedings of the Workshop on Visual Motion*, pages 236–241, Irvine, California, March 1989.

[43] S.-L. Peng, *Image Sequence Analysis Using Temporal Slices*, Ph.D. thesis, University of Southern California, January 1991, IRIS Technical Report 265.

[44] V. K. Prasanna Kumar, *Parallel Algorithms and Architectures for Image Understanding*, Academic Press, Boston, 1991.

[45] A. S. Rao and V. K. Prasanna, "Efficient Parallel Algorithms for *k*-dimensional Images," In *International Conference on Image Analysis and Processing*, 1991.

[46] K. Rao, *Shape Description from Sparse and Imperfect Data*, Ph.D. thesis, University of Southern California, December 1988, IRIS Technical Report 250.

[47] K. Rao and G. Medioni, "Useful Geometric Properties of the Generalized Cone," In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 276-281, 1988.

[48] K. Rao and R. Nevatia, "Description of Complex Objects from Incomplete and Imperfect Data," In *Proceedings of the Image Understanding Workshop*, pages 399-414, Palo Alto, California, May 1989.

[49] C. Reinhart and R. Nevatia, "Efficient Parallel Processing in High Level Vision," In *Proceedings of the DARPA Image Understanding Workshop*, pages 829–839, Pittsburgh, PA, September 1990.

[50] C. Reinhart and R. Nevatia, "Parallel Linear Feature Extraction," In *Proceedings of the DARPA Image Understanding Workshop*, San Diego, CA, January 1992.

[51] C. Reinhart, *Specifying Parallel Processor Architectures for High-Level Computer Vision Algorithms*, Ph.D. thesis, University of Southern California, 1991.

[52] H. Rom and G. Medioni, "Hierarchical Decomposition and Axial Representation of Shape," In *Proceedings of the DARPA Image Understanding Workshop*, San-Diego, California, January 1992.

[53] H. Rom and G. Medioni, "Hierarchical Decomposition and Axial Shape Description," In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 49–55, Champaign, Illinois, June 1992.

[54] H. Rom and G. Medioni. "Hierarchical Decomposition and Axial Shape Description," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):973–981, 1993.

[55] P. Saint-Marc, H. Rom, and G. Medioni, "B-Spline Contour Representation and Symmetry Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10), 1993.

[56] F. Stein and G. Medioni, "Efficient Two Dimensional Object Recognition," In *Proceedings of International Conference on Pattern Recognition*, Atlantic City, New Jersey, June 1990.

[57] F. Stein and G. Medioni, "Structural Hashing: Efficient Three Dimensional Object Recognition," In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 244–250, Maui, Hawaii, June 1991.

[58] F. Stein and G. Medioni, "Structural Indexing: Efficient Three Dimensional Object Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 14*, No. 2, pages 125–146, February 1992.

[59] F. Stein and G. Medioni, "Recognition of 3-D Objects from 2-D Groupings," *Proceedings of the DARPA Image Understanding Workshop*, January 1992, pages 667-674.

[60] F. Ulupınar and R. Nevatia, "Using Symmetries for Analysis of Shape from Contour," In *Proceedings of the IEEE International Conference on Computer Vision*, pages 414–427, Tampa, Florida, December 1988.

[61] F. Ulupınar and R. Nevatia, "Inferring Shape from Contour for Curved Surfaces," In *Proceedings of the International Conference on Pattern Recognition*, volume 1, pages 147–154, Atlantic City, New Jersey, June 1990.

[62] F. Ulupınar and R. Nevatia, "Shape from Contour: SHGCs," In *Proceedings of the IEEE International Conference on Computer Vision*, pages 582–582, Osaka, Japan, December 1990.

[63] F. Ulupınar and R. Nevatia, "Recovering Shape from Contour for Constant Cross Section Generalized Cylinders," In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 674–676, Maui, Hawaii, June 1991.

[64] F. Ulupınar and R. Nevatia, "Recovery of 3-D Objects with Multiple Curved Surfaces from 2-D Contours," In *Proceedings of the DARPA Image Understanding Workshop*, San Diego, CA, January 1992.

[65] F. Ulupinar, *Perception of 3-D Shape from 2-D Image of Contours*, Ph.D. Thesis, Computer Science, University of Southern California, Los Angeles, August, 1991.

[66] M. Zerroug and R. Nevatia, "Scene Segmentation and Volumetric Descriptions of SHGCs from a Single Intensity Image," In *Proceedings of the Image Understanding Workshop*, pages 905-916, Washington DC., 1993.

[67] M. Zerroug and R. Nevatia, "Quasi-Invariant Properties and 3-D Shape Recovery of Non-Straight, Non-Constant Generalized Cylinders," In *Proceedings of the DARPA Image Understanding Workshop*, Washington, DC, April 1993.

[68] M. Zerroug and R. Nevatia, "Quasi-invariant Properties and 3-D Shape Recovery of Non-Straight, Non-Constant Generalized Cylinders", In *Proceedings of Computer Vision and Pattern Recognition*, pages 96-103. 1993. New York.

[69] M. Zerroug and R. Nevatia. "Volumetric Descriptions from a Single Intensity Image," *International Journal of Computer Vision*, 1993, (To appear).

5.2 Other References

[70] P. J. Besl and R. C. Jain. "Invariant Surface Characteristics for 3D Object Recognition in Range Image," *Computer Vision, Graphics and Image Processing*, 33:33–80, 1986.

[71] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion," *International Journal of Computer Vision*, 1:7–55, 1987.

[72] V. Guillemin and A. Pollack, *Differential Topology*, Prentice Hall, 1974.

[73] R. I. Hartley, "Calibration of Cameras Using the Essential Matrix," in *Proceedings of the DARPA Image Understanding Workshop*, pages 911-915, San Diego, CA, January 1992.

[74] B. K. P. Horn, "Relative Orientation Revisited," in *Journal of the Optical Society of America A., Vol. 8*, NO 10, pages 1630-1638, October 1991.

[75] J. J. Koenderink, *Solid Shape*, MIT Press, 1990.

[76] H.C. Longuet Higgins, "*A* Computer Algorithm for Reconstructing a Scene from Two Projections," in *Nature, Vol. 293*, pages 133-135, September 1981.

[77] R. MacGregor and M. Burstein, "Using a Description Classifier to Enhance Knowledge Representation," *IEEE Expert, Vol 6*, No. 3, pages 41-46, June 1991

[78] R. Miller and Q. F. Stout, "Geometric Algorithms for Digitized Pictures on a Mesh-Connected Computer," In *IEEE Transactions on Pattern. Analysis and Machine Intelligence*, 1985.

[79] L. R. Nackman and S. M. Pizer, "Three-Dimensional Shape Description Using the Symmetric Axis Transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):187–202, March 1985.

[80] J. Ponce and M. Brady. "Towards a Surface Primal Sketch," In T. Kanade, editor, *Three Dimensional Machine Vision*, pages 195–239. Kluwer Academic, New York, 1987.

[81] J. Ponce, D. Chelberg, and W. B. Mann, "Invariant Properties of Straight Homogeneous Generalized Cylinders and Their Contours," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(9):951–966, 1989.

[82] S.A. Shafer and T. Kanade, "The Theory of Straight Homogeneous Generalized Cylinders," *Technical Report* CS-083-105, Carnegie Mellon University, 1983.

[83] C. Weems, S. Levitan, A. Hanson, and E. Riseman, "The Image Understanding Architecture," *International Journal of Computer Vision*, 2(3):251–282, January 1989.