ETL-0580

Robust Image Understanding: Techniques and Applications

First Annual Report

Azriel Rosenfeld

University of Maryland
Center for Automation Research
College Park, MD 20742-3411

December 1990

Prepared for:
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209-2308

U.S. Army Corps of Engineers
Engineer Topographic Laboratories
Fort Belvoir, Virginia 22060-5546

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1990	Annual September 1989 - July 1990

4. TITLE AND SUBTITLE

Robust Image Understanding - Techniques and Applications
First Annual Report

5. FUNDING NUMBERS

DACA76-89-C-0019

6. AUTHOR(S)

Azriel Rosenfeld (compiler)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Maryland
Center for Automation Research
College Park, MD 20742-3411

8. PERFORMING ORGANIZATION REPORT NUMBER

Abstract 1096

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

US Army Engineer Topographic Laboratories
Fort Belvoir, VA 22060-5546

Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209-2308

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ETL-0580

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The research on the contract dealt with image understanding applications to both navigation and recognition. Thirteen technical reports were issued on the contract during this period referred to by numbers in brackets in the remainder of the report.

Research on navigation was concerned with the following specific topics which are discussed in further detail in this report:

a. analysis of superimposed moving patterns [1,2], b. path and motion planning [5,13], c. structure from motion [6], d. motion uncertainty [8], e. motion illusions [11], and f. motion recovery in the presence of discontinuities [12]. Recognition research was concerned with: g. recognition of compact shapes by energy function minimization [3], h. learning of invariant shape properties [4], i. slant-insensitive shape descriptors [7], and j. edge detection [9] and line fitting [10].

14. SUBJECT TERMS

image understanding, vision, path planning, navigation, edge detection

15. NUMBER OF PAGES

22

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	UNLIMITED

PREFACE

This report describes work performed under Contract DACA76-89-C-0019 by the Center for Automation Research, University of Maryland, College Park, Maryland for the U.S. Army Engineer Topographic Laboratories (ETL), Fort Belvoir, Virginia, and the Defense Advanced Research Projects Agency (DARPA), Arlington, Virginia. The Contracting Officer's Representative at ETL is Ms. Rosalene Holecheck. The DARPA points of contact are Dr. Erik Mettala and Dr. Rand Waltzman.

Contents

1. Introduction

Contract DACA76-89-C-0019, for research on Robust Image Understanding—Techniques and Applications, was awarded to the University of Maryland on 28 September, 1989. This report describes the work done on the contract during the period September 1989 – July 1990.

The research on the contract dealt with image understanding applications to both navigation and recognition. Thirteen technical reports were issued on the contract during this period; they are listed in Section 12, and are referred to by numbers in brackets in the remainder of this report.

Research on navigation was concerned with the following specific topics:

a) Analysis of superimposed moving patterns [1,2]

b) Path and motion planning [5,13]

c) Structure from motion [6]

d) Motion uncertainty [8]

e) Motion illusions [11]

f) Motion recovery in the presence of discontinuities [12]

Recognition research was concerned with

g) Recognition of compact shapes by energy function minimization [3]

h) Learning of invariant shape properties [4]

i) Slant-insensitive shape descriptors [7]

j) Edge detection [9] and line fitting [10]

These topics are discussed in the following sections of this report.

2. Analysis of superimposed moving patterns

As we move relative to trees and bushes in a forest we can perceive that their leaves and branches move in depth. One of the primary sources which permits us to segment the trees and bushes into regions of different depth is given by the optical flow field. Under appropriate conditions there exists an intrinsic relation between the motion of objects in space and the optical flow. For example, the relative motion between an observer and rigid objects with smooth surfaces generates regions of smoothly varying optical flow in the image plane. Current motion theories used for the reconstruction of the optical flow field assume prior information on the optical flow in order to make the problem well-posed from the point of view of regularization theory. These priors enforce a smoothness constraint on the optical flow field, and this is only consistent with the motion of objects with smooth surfaces; in this case the flow field can only assume one value at each image point. In the presence of motion discontinuities or transparent superimposed surfaces in relative motion the optical flow can have more than one value at each image point, and this is not addressed by current motion theories. We have developed a statistical model for the analysis of superimposed patterns in relative motion.

The perception of superimposed patterns moving independently relative to one another was first studied by Adelson and Movshon. By using two sinusoidally modulated contrast functions in relative motion, with known velocity, contrast, and spatial frequency, they performed psychophysical experiments in order to determine the conditions under which we can see motion transparency or coherence. Motion coherence corresponds to the perception of a compound pattern moving with a single velocity in a given direction; in the case of motion transparency the two patterns are seen as moving independently one across the other. They showed that, for fixed velocities and contrast, we perceive motion transparency if the difference between their spatial frequencies is high. On the other hand, if both patterns have similar spatial frequencies, contrasts, and speeds, then we perceive motion coherence. More recently Stoner, Albright, and Ramachandran studied the conditions for the perception of motion coherence or transparency for superimposed transparent patterns. They showed that the probability of perceiving motion coherence or transparency depends on the luminance of

the regions of pattern intersection.

Previous studies of motion transparency and coherence have dealt with superimposed sinusoidal bar patterns whose gray levels combine additively; but the "transparent" patterns encountered in the real world are composed of opaque elements with gaps between them. We have analyzed motion transparency for such binary patterns; for simplicity, we have used patterns made up of lines. The case of straight line patterns is treated in [1], and that of curvilinear patterns in [2].

Specifically, we have analyzed cases involving two superimposed line patterns moving in the frontoparallel plane. If these patterns have regions of high curvature, or features like end-points or corners, we are able to solve the aperture problem for each pattern separately, and consequently we perceive motion transparency. On the other hand, in the absence of features, or for small curvature, we perceive motion coherence which is given by the motion of the compound pattern. We have developed a statistical model for the perception of motion transparency and coherence which is given by a two-stage process for the extraction of the optical flow and the velocity histogram. The velocity histogram, which is a plot of the number of occurrences of each velocity vector, is unimodal for motion coherence and bimodal for motion transparency. The image is divided into regions, and inside each of them we compute the optical flow; for line end-points and corners we compute their velocities by matching them between images; and for lines we combine the normal velocity components by computing the intersection of the corresponding constraint lines in the velocity space. We use a generalized version of the two-stage process for the extraction of the optical flow which takes into account superimposed patterns. Our model is also able to predict the transition between the perception of motion transparency and coherence, and it is in good agreement with informal perceptual experiments done with line patterns.

3. Path and motion planning

The problem of navigation is related to sensory mediated movement and usually refers to the ability to move successfully from one point to another in some environment. The problem, for the most part, has been treated in the robotics literature from the viewpoint of path

planning. In this work, complete knowledge of the environment is assumed. But in reality, information about the environment is generally unknown (with the exception of some specific industrial environments) and it is acquired through some sensory modality—for example, touch (or hitting an obstacle) has been used in a static unknown environment for the purpose of planning a successful path. In our research the sensory modality used to acquire environmental information is vision. Using as input a series of images acquired by the navigating robot, we find how the robot should move in order to accomplish its task (i.e. plan a successful (safe) path). Given the nature of the problem (incomplete information), we cannot guarantee that the robot will accomplish the task.

We have developed [5] a solution to the problem of finding the 3-D motion of a moving obstacle on the basis of visual information. We assume that the obstacles are translating locally and that they are sufficiently textured so that there are enough "points of interest" on their surfaces (and consequently on their images). Our imaging system consists of four cameras and we assume that the obstacle in motion is visible from all four cameras. Each obstacle, for the vision algorithm, is treated as a cloud of points in 3-D; these points are extracted through some interest point operator. The algorithm does not use any point correspondences. It derives the 3-D motions of the obstacles directly from the image data. The solution has been shown to be robust.

Using this information, we can treat the path planning problem as one of moving a robot (assumed to be a polygon) from an initial point to a goal point in the presence of other moving polygons whose motions are uncertain. We assume that there are bounds on the speed and acceleration of the robot. Under these assumptions we want to optimize the probability of reaching the goal.

Because of the uncertainty of the obstacles' motion, we regard them as growing in space-time. We can then determine what regions are reachable by the robot without changing velocity in a given interval. We then need to determine trajectories that reach the goal. We have considered two specific problems:

(i) finding smooth acceleration trajectories, and

(ii) obtaining strategies that try to minimize the maximum acceleration.

Our solutions are useful contributions to the problem of planning paths using local inaccurate information. The details can be found in [5].

In more recent work [13] we have addressed the problem of efficiently planning a path for a robot between two points when the path is forced to change dynamically by the occurrence of certain *events* in the environment. The *event*, for example, may be the discovery of another moving object on a collision course with the robot. The robot would be forced to take evasive action whenever such an *alarm* occurs. We have developed a probabilistic model that represents the dynamic behavior in terms of *alarms* following a Poisson distribution, and *safety rules* that assume that some regions are *safe*. We have derived a provably optimal expected solution for the problem and studied the effect of the probabilistic parameter (λ) of the dynamic environment on the optimal path, and the effect of "vision" (or *time to collision*) on the planned paths. The results can be used in designing heuristics for path planning in a more general framework. Our study gives insights into the role of various parameters on the *average* efficiency of path-planning in a simple dynamic, unknown environment. The simplicity of the model used in justified in the difficulty of analyzing a more complicated (unknown) dynamic environment, and by the generality of the results obtained using this simple model.

4. Structure from motion

When one looks at a real image one can see a number of identifiable features such as points and lines. It has been argued that these features are the only things needed for computing the motion and structure because they carry reliable information, there exist mathematical and computational tools to treat them, and extensive experience and literature from photogrammetry can be tapped. On the other hand the number of image pixels that are covered by these features is only a tiny fraction of the total number of pixels in the image. If one uses only point and line features, the vast majority of the image remains unused. The feature points carry more information than the rest of the pixels, but the rest of the pixels are much more numerous. They should not be left unused. This underutilization is not the only problem; two additional, more solid problems arise. First, there is no consensus among

researchers in computer vision on what a feature point or line is, either in a rigorous mathematical sense or even in an intuitive practical sense, judging from what different detectors detect as features. The main consequence of this is that there is no general algorithm to detect features and match them. Second, even if one can detect point and line features, there exists no algorithm that works with both of them at the same time and guarantees a unique solution, although there are algorithms for each one separately.

There is another approach to structure from motion that assumes continuous motion and grey level images that are differentiable in time and space. Despite its theoretical elegance, this approach is plagued by the aperture problem: the motion (optic flow) of a point on a moving curved line (isophote, zero crossing, etc.) cannot be recovered fully; we can recover only its projection on the normal to the line. Most algorithms based on this approach assume that derivatives up to second order of the (not fully known) image flow are given or can be computed, and that the flow is smooth. The resulting algorithms are local (and hence unstable) in nature.

Obviously there is need to overcome these difficulties and combine the advantages of all the existing methods. This does not seem to be an easy task within the existing theories, since they all are rather incompatible, with different input requirements and conflicting assumptions, and they operate on different geometric entities. But are points, lines, curves or isophotes different entities or can they be defined in a uniform way? The truth is that they are different when compared as abstract geometric entities. In the context of visual motion a feature point is a small area of the image that (besides the statistics of its grey level that made the detector locate it) is moving with a motion whose uncertainty is more or less circularly symmetric and small (finite). In the same context, a line is an area whose motion has an uncertainty that is finite in one direction (normal to the line) and infinite in the other direction (along the line). These definitions, having an obvious statistical flavor, seem natural and uniform. We can forget the separate definitions of points and lines and concentrate instead on the statistics of the disparity fields. The only geometric entity we need is the curve in its most general sense, whether it be a chain of edgels, a zero crossing, or an isophote. Thus any point in the image can be considered. If such a curve is moving then a point that was on it in the first frame will move with the curve but we do not know

where on the length of the curve it will be in the next frame. We only know that it is on the curve but we have only a probability distribution for where it can be on the curve. This probability distribution accounts for the tangential component of the uncertainty of the motion of the point. There is also another component of uncertainty, the normal one, because due to the fuzziness of the curve we cannot assume that the motion of the point along the normal direction can be recovered exactly. Thus we have a probability distribution along both directions. So we do not assume that we shall be given the exact displacement vector of any point but only a probability distribution for it. As a working assumption we assume that this distribution is Gaussian and its parameters are given. Gaussian is a very good choice because it makes sense intuitively and leads to very stable statistics, meaning that if this assumption holds only approximately then the consequences are not catastrophic and the degradation is graceful. Also, coupled with maximum likelihood it gives rise to least-squares estimators which have nice analytic expressions [6].

Part of the difficulty in structure from motion is its absolute separation from the preceding stage of computing the displacement vector field (or flow field or correspondence; they all are of the same flavor). This unavoidably leads to the idea of trying to find the exact displacement field using restrictive assumptions such as smoothness and then, pretending that this is the correct field, find the structure and motion. In our approach we require much less from the preceding step than a complete, accurate disparity field, thus eliminating the complete reliance on assumptions such as smoothness. Then, using only the assumptions of rigidity and Gaussian noise, we find the motion and structure (with some uncertainty depending on the data). If indeed the error in the data follows the Gaussian distribution, then the structure and motion we compute is the optimal one. Otherwise it is not, and what we have computed is a good approximation which is the collective result of information from a large area of the image. Once we have this collective result, which is a constraint on the flow (by backprojection), we can couple it with the original grey level based constraint and compute the displacement field again.

5. Motion uncertainty

In the process of extracting velocity through energy filters, it is possible to reduce the motion uncertainty to a minimum if we choose the spatial and temporal filter sizes such that the temporal bandwidth is larger than the spatial bandwidth. This means that it is possible to reduce the arbitrariness in the choice of filter parameters if we assume that the error in the extraction of velocity is a minimum. In our analysis [8], we assumed that the images are highly textured, so that they have a flat power spectrum. The image noise was assumed to be white and Gaussian, which leads to a χ^2 energy noise probability distribution.

The process of extracting velocity (optical flow) by space-time filtering uses the convolution of a sequence of images with a collection of space-time filters each of which is tuned to a given space-time orientation. The individual energy filters are not velocity tuned, and therefore it is necessary to use a collection of them in order to extract velocity. One of the consequences of this is that the extraction of velocity through these filters can only be done with limited precision, and so there always exists a non-zero motion uncertainty.

We have analyzed the properties of space-time energy filters, and in particular we have computed and analyzed the motion uncertainty An elegant way to study the properties of the motion uncertainty is through the use of the Cramér-Rao inequality which gives us the lower bound on the mean square error for the estimation of the velocity. To deduce this inequality we need to describe the process of extracting velocity in the framework of estimation theory, and this requires the knowledge of a conditional probability function. By assuming that the images are corrupted by white Gaussian additive noise we were able to show that the conditional probability function for the energy noise is given by the χ^2 distribution. The resulting motion uncertainty lower bound depends on the velocity and on the filter parameters. We have shown that this lower bound is minimum, that is, it is, on average, smallest and smoothest as a function of the velocity, when the filter parameters vary inside a given range of values. In particular we have explicitly computed the motion uncertainty lower bound for the Gabor energy filter and have shown that this lower bound is minimum if the temporal bandwidth of the filter is larger than the spatial bandwidth.

The fact that the temporal bandwidth is larger than the spatial bandwidth is in accord

with physiological data obtained from the primary cortex of cats. We may therefore conjecture that this difference between the spatial and temporal bandwidths has the purpose of making the extraction of velocity through cells in the primary visual cortex more efficient, that is, it minimizes the motion uncertainty. Here we are not interested in describing a detailed model for motion sensitive cells in the primary visual cortex, but instead in *qualitatively* analyzing some common properties of these cells and energy filters. We conjecture that, because the motion uncertainty lower bound is minimum for the case in which the temporal bandwidth of the Gabor energy filter is larger than its spatial bandwidth, motion sensitive cells are built in such a way that they extract velocity in an *efficient* form.

There are some important open questions which are worth mentioning. First, consider the sampling of the energy filter orientation. Instead of arbitrarily fixing the frequency orientations we could assume that the energy filters "learn" these orientations from examples. The most important point to be proven here is that we should get qualitatively similar values for the minimum uncertainty as the number of samples becomes large. Second, we should be able to obtain similar results by using other oriented filters. Third, for the case of having to deal with an arbitrary type of image model, which is different from the flat power spectrum images used here, it is necessary to deal with the aperture problem, which, in the framework of energy filters, is still unresolved. Fourth, there exists a collection of open problems in connection with the mathematical modelling of motion sensitive cells in biological systems. As one example we mention the need to use more realistic oriented filters: the Gabor filter, although simple in its mathematical structure, is non-causal in the space-time domain.

6. Motion illusions

Humans use various cues in order to understand the structure of the world from images. One such cue is the contours of an object formed by occlusion or from surface discontinuities. It is known that contours in the image of an object provide various amounts of information about the shape of the object in view, depending on assumptions that the observer makes. Another powerful cue is motion. The ability of the human visual system to discern structure from a motion stimulus is well known and it has a solid theoretical and experimental foundation.

But when humans interpret a visual scene they use various cues in order to understand what they observe, and the interpretation comes from combining the information acquired from the various modules devoted to specific cues. In such an integration of modules it seems that each cue carries a different weight and importance.

We have performed several experiments [11] in which we made sure that the only cues available to the observer were contour and motion. It turns out that when humans combine information from contour and motion to reconstruct the shape of an object in view, if the results of the two modules—shape from contour and structure from motion—are inconsistent, they totally discard one of the cues and an illusion is experienced.

The mathematical model that we have introduced reconstructs, from occluding contour and local motion information, the surface which is as smooth as possible and best satisfies the motion constraint everywhere, while satisfying the boundary conditions provided by the form of the contour. The results of applying the algorithm to a wide variety of illusion-producing inputs are also consistent with human perception. It thus appears that when humans combine the cues of occluding contours and local motion to reconstruct the moving object in view, they reconstruct (see) the surface which is consistent with the boundaries, is as smooth as possible, and best satisfies the local motion constraints.

Based on our experimental and theoretical results, one cannot reject the hypothesis that cortical connections in the primate visual cortex implement some form of regularization for motion perception, in both the estimation of retinal motion and its interpretation for the purposes of reconstruction. It has been suggested that the theory of regularization may be used as a theory for low-level vision. Our work demonstrates that it is beneficial to think in this paradigm.

7. Motion recovery in the presence of discontinuities

Many problems in low-level vision are ill-posed in the sense that their solutions do not exist, are not unique, or do not vary continuously with the data. We are primarily concerned with the difficulties caused by substantial amounts of noise or inaccurate constraints relating the image data to the unknowns. We also often have more unknowns than constraints, so there

is no unique solution. We need to use a priori information about smoothness (regularization) in order to handle these difficulties, but we do not want to smooth over discontinuities. We must also realize that we do not know a priori the relative amounts of smoothness and noise or the exact probability distributions of measures of noise and smoothness. A well-known paradigm of discontinuous regularization requires solution of variational conditions with multiple local minima and is not maximally robust against the possibility that we have misspecified the probability distribution of the smoothness measure. The condition must be solved by slow Monte-Carlo methods, by deterministic methods that may not find the optimal solution, or by continuation methods that if properly implemented always work but are sometimes very slow.

We have applied [12] Huber's theory of robust statistics (so-called M-statistics) to obtain a convex variational condition which has a unique solution. The problem we have addressed is optical flow or more precisely small-motion depth from stereo. Thus we know the motion is translational. Because of calibration errors, we do not know the exact direction or distance between the two cameras. Ignorance of the distance means that, unavoidably, our depth estimates are only accurate modulo a scale factor that we cannot precisely estimate. We can use image data to refine our rough knowledge of inter-camera directions. Our primary object is to determine the depth at each point.

Stereo is a typical example of an ill-posed problem. We are interested in the case of short-range motion. Thus we really are estimating flow. The standard optical flow equation is extremely noisy, so it is important to use a priori smoothness information to regularize the solution, but we do not wish to smooth over discontinuities. Thus we should penalize deviations from smoothness but we should not over-penalize large deviations. We do not know how to penalize such deviations; so we use the penalty function that works best in the worst case. This function is convex and thus we do not need to worry about multiple local minima. We are not implicitly assuming the scene in view to be piecewise smooth modulo Gaussian fractal deviations, so we can handle deviations from smoothness such as rounded corners. This method of convex or robust regularization can also be applied to smoothing the intensity function which we need to do in order to estimate the derivatives of intensity occurring in the flow equation.

The flow equation itself, however, is very unreliable at certain points and we can use convex regularization to study errors in the flow equation caused by errors in the fundamental assumption that corresponding points have the same intensity. Thus the error term tends to be smooth and small, but not everywhere is it smooth and small.

The unreliability of the flow equation at certain points is a fundamental issue in motion research; as we have seen, this can be handled to some extent using convex regularization. But there are some difficulties that still have to be confronted; we have been emphasizing the errors in the flow equation due to corresponding points not having the same intensity, but there is also an error that occurs because it is intrinsically difficult to compute derivatives. The error due to derivative misestimation varies greatly from point to point. We would like the relative weight of the flow term λ in the variational condition to reflect this variability; thus λ should vary with position. It is not obvious how to do this: natural suggestions one could make are that λ should depend on the second derivatives of intensity or on the sum of the data consistency and smoothness terms in the variational condition that we use to smooth the intensity. In both cases, we are using a crude estimate of how much the derivatives vary from pont to point in a neighborhood of the given point. The danger is that we will down-weight points with large intensity derivatives and it is at these points that depth discontinuities tend to occur. This is a topic for future research.

8. Recognition of compact shapes

We have developed a method of recognizing compact objects in an image by energy function minimization. The energy function is based on a polar coordinate object representation, defined using any center from which the object's contour is visible. It incorporates both low-level and high-level information about the object: contour sharpness and smoothness at the low level, and contour shape at the high level.

Our approach [3] differs from previous work on optimization-based methods for shape extraction in several important respects:

1. It represents the object in polar form; this constrains the contour of the object to surround the center and to be entirely visible from it.

2. It incorporates both local information (contour sharpness and smoothness) and global information (object shape) simultaneously during the optimization, and operates on raw image data, thus making backtracking unnecessary.

3. These types of information are given weights that change during the course of the optimization.

4. No noise model assumptions are needed. Conditional probabilities are not calculated directly.

5. The method is highly parallelizable, since it makes use only of local information.

6. Highly correlated geometrical models can be employed, thus enabling the discrimination of subtly different objects.

7. The algorithm outputs the object's classification along with a measure of confidence. Unknown objects can be identified as such.

8. The complexity of the energy function does not greatly increase the computational burden on the procedure. The increase is on the order of a constant.

We define a compact object in an image as an object that can be represented in polar coordinate form by a smoothly varying radius function. Our basic approach consists of three main stages:

1) **Detection**. Detecting a candidate object "centered" in the image. This center can be anywhere inside the object, preferably close to its centroid.

2) **Representation**. Representing the candidate object in polar coordinate form, relative to the center, and identifying the location of its boundary along each radius.

3) **Matching**. Comparing the polar coordinate representation of the object with a set of stored representations.

We have incorporated the second and third stages into the energy function. This has the advantage that all known information about the objects is utilized simultaneously. Our use

of the polar coordinate representation reduces the dimension of the Markov random field and increases the efficiency of the solution. More specifically,

a) It reduces the optimization problem from 2-D to 1-D, resulting in a considerable reduction in the amount of computation needed.

b) It uses the image grey level information as data in an optimization process bent on detecting the best field configuration, which is given in distance units.

c) The radial scheme provides a scaling and orientation invariant representation which is highly compatible with compact object recognition.

The energy function used is also a novel feature of this approach. Two weight functions W_1 and W_2 direct the optimization by putting the emphasis on the proper energy function level. Initially the low level is dominant and the high level is used to keep the configuration compact in a very general way. Later, as W_2 increases in value as the confidence of the match improves, the low level performs the function of insuring that the high level does not deviate from what is present in the raw data.

9. Learning of invariant shape properties

We receive knowledge form the world around us through our various senses: that is, each bit of information from our senses corresponds to an *input* from the outside world to the brain. For example, the retina of the eye may be likened to a large number of binary inputs. A particular input would have value 1 if the rod or cone retinal cell that it corresponds to were currently picking up a particular kind of signal. Correspondingly, an output would have value 1 if a particular target object were present in the visual field.

Many of the objects recognized by the brain correspond to certain combinations of the states of these inputs. These objects can be very simple and well-defined (such as "square") or complicated and harder to define precisely (such as "chair"). In this context *learning* an object may be defined as finding those input combinations that correspond to the target object. This means determining (1) which of the many inputs are relevant to the existence of

that concept and (2) the states (1 or 0) of those relevant binary inputs. Using this knowledge, the brain can create a *detector* for that object. If those relevant inputs are all in the correct states, then the detector will indicate that the target object exists in the sensory field.

Point (1), the isolation of relevant input, is an important and often overlooked part of the learning process. Supposed we wish to learn the concept of "square"—that is, to form a detector that determines whether or not a square is formed by a set of active input (retinal) units. In this case, we are only interested in the square itself, and not in any other information available in this context. This detector checks those and only those inputs necessary to confirm the square. Other inputs corresponding to "background noise" are ignored. If these inputs were not ignored, then the brain might have to construct an enormous number of square detectors, one for each different kind of "background noise". In previous work we have outlined a system called *constraint motion* that is capable of this kind of learning.

So far, we have described target objects in terms of a set of *fixed* inputs. If these specific inputs are in the correct state, then the object is present in the image; otherwise, it is not. In most cases, however, an object is not restricted to a single area of the visual field. We can easily recognize objects independent of conditions such as translation, rotation, or scale change in the image. That is, object detection is *invariant* with respect to these conditions.

We have expanded the idea of constraint motion learning [4] to take into account the problems associated with invariance. One of the strengths of the constraint motion system is its ability to learn many concepts simultaneously. Each detector starts with an example of a specific concept, and over time it comes to focus on that concept, learning the relevant inputs while ignoring data from other concepts being learned by other detectors.

We have shown that the constraint motion system can use a similar method to focus on the most likely transformations of the target object, giving it a better idea of its relevant inputs. We have mathematically described the learnability of a concept as a function of the number of possible transformations.

An interesting result of our work is the similarity we have found between the credit assignment problem for learning with multiple processors and the problem of learning the location of an unknown object in an image. The same properties of the constraint motion algorithm that facilitate distributed learning also expedite learning in the presence of invariance.

10. Slant-insensitive shape descriptors

Recognizing a 3-D shape on the basis of its single 2-D image (from an unknown viewpoint) is difficult because the shape looks different from different viewpoints and is partly visible from any viewpoint. When a 3-D object rotates, the transformation of its 2-D silhouette depends on its shape. If the object is flat, the transformation is equivalent to perspective distortion; if it is solid, parts of the silhouette can be distorted to a degree greater or less than perspective distortion.

We have developed a method of shape recognition based on analysis of the shape's silhouette, using descriptors that are insensitive to perspective distortion (i.e., to slant), and that also allow the slant of the shape to be estimated. The details of the method and examples of its performance can be found in [7].

11. Edge detection and line fitting

An edge, a discontinuity, or an abrupt change in gray level or color, is one of the fundamentally important primitive features of an image necessary for image analysis. Edge detection, a local operation at every pixel, can be classified into two categories: template matching and discrete approximations of differential operators. An important set of template-matching operators are the Frei-Chen masks. These 3×3 masks were proposed on the basis of a vector space approach, but the way the masks were chosen was not fully explained.

We have developed [9] an interpretation of the Frei-Chen masks in terms of eight-dimensional Fourier transform coefficient vectors. The linear transformation between nine-dimensional Frei-Chen space and the eight-dimensional Fourier transform has been derived. We have also proposed a modified set of eight orthogonal masks based on the frequency space analysis.

Detection of straight edges is usually based on fitting straight lines to detected edge points. A set of n distinct points in the plane defines $\binom{n}{2}$ lines by joining each pair of distinct points. The median slope of these $O(n^2)$ lines was proposed by Theil as a robust estimator for the slope of the line of best fit for the points.

We have developed [10] a randomized algorithm for selecting the k-th smallest slope of such a set of lines which runs in expected $O(n \log n)$ time. Our emphasis has been on designing an algorithm which is provably efficient (with very high probability), which handles degenerate cases correctly, and which has a simple and efficient implementation. We have experimented extensively with the implementation in order to establish its efficiency and robustness.

12. References

[1] Azriel Rosenfeld, Radu S. Jasinschi, and Helder J. Araujo, "Motion Transparency I: Straight Line Patterns", CAR-TR-465, CS-TR-2325, September 1989.

[2] Radu S. Jasinschi, Azriel Rosenfeld, and Helder J. Araujo, "Motion Transparency II: Curved Line Patterns", CAR-TR-477, CS-TR-2368, December 1989.

[3] N.S. Friedland and Azriel Rosenfeld, "Compact Object Recognition Using Energy Function Based Optimization", CAR-TR-478, CS-TR-2369, December 1989.

[4] John R. Sullins, "Distributed Learning under Invariance", CAR-TR-479, CS-TR-2370, December 1989.

[5] Anup Basu, "A Framework for Motion Planning in the Presence of Moving Obstacles", CAR-TR-481, CS-TR-2378, December 1989.

[6] Minas Spetsakis and John (Yiannis) Aloimonos, "Unification Theory of Structure from Motion", CAR-TR-482, CS-TR-2379, December 1989.

[7] Zygmunt Pizlo and Azriel Rosenfeld, "Recognition of 2-D Shape in 3-D Space", CAR-TR-484, CS-TR-2381, January 1990.

[8] Radu S. Jasinschi, "Energy Filters, Motion Uncertainty, and Motion Sensitive Cells in the Visual Cortex: A Mathematical Analysis", CAR-TR-487, CS-TR-2397, February 1990.

[9] Rae-Hong Park, "A Fourier Interpretation of the Frei-Chen Edge Masks", CAR-TR-489, CS-TR-2414, February 1990.

[10] Michael B. Dillencourt, David M. Mount and Nathan S. Netanyahu, "A Randomized Algorithm for Slope Selection", CAR-TR-493, CS-TR-2431, March 1990.

[11] John (Yiannis) Aloimonos and Liuqing Huang, "Motion-Boundary Illusions and their Regularization", CAR-TR-495, CS-TR-2434, March 1990.

[12] David Shulman and Jean-Yves Hervé, "On the Regularization of Discontinuous Flow Fields", CAR-TR-498, CS-TR-2440, March 1990.

[13] Rajeev Sharma, "Locally Efficient Path Planning in a Dynamic Environment with a Probabilistic Model", CAR-TR-507, CS-TR-2488, May 1990.