SRI International

PROSPECTS FOR INDUSTRIAL VISION

Technical Note 175

November 1978

By: Jay M. Tenenbaum, Program Manager
Harry G. Barrow, Senior Computer Scientist
Robert C. Bolles, Computer Scientist

Artificial Intelligence Center

Paper presented at General Motors Symposium on Computer Vision and Sensor-Based Robots, Warren, Michigan, September 1978. (Proceedings to be published by Plennum Press.)

SRI International

333 Ravenswood Ave. • Menlo Park, California 94025
(415) 326-6200 • Cable: SRI INTL MPK • TWX: 910-373-1246

I. Overview of the Paper

Industrial Vision Requirements
Limitations of Current Systems
Need for General-purpose Vision
Proposed System Design
Research Directions

The proceedings of this symposium contain many impressive examples of what can be accomplished today in factories with computer vision. Compared with human vision, however, the capabilities of these specialized systems are still very rudimentary. In this paper, we will examine some important industrial vision requirements that are simple for humans to fulfill but well beyond the ability of any existing machine. We will try to understand what limitations of current systems make these tasks so hard for a machine and use this knowledge to outline the design of a general-purpose, computer-vision system, capable of high performance in a wide variety of industrial vision tasks. We will conclude by suggesting some promising research directions towards realizing such a system.

II. Performance Characteristics

Cost
Speed
Competence
Reliability
Flexibility
Trainability

The performance of a vision system can be evaluated along numerous dimensions. As researchers, we view competence as the most fundamental since it determines whether a task can be performed at all. If the basic competence exists, then it is up to the ingenuity of industrial engineers to bring factors such as speed, cost, and reliability within acceptable limits so that an application becomes economically cost-effective.

III. Limitations of Current Industrial Vision Systems

High Contrast
No Shadows
No Occlusion
2-D Models
Rigid Objects
Standard Viewpoint

The competence of current industrial vision systems restricts their application, by and large, to situations where individual objects can be easily isolated in an image. Typically, objects are presented against a high contrast background (no occlusion) with lighting controlled to eliminate shadows, highlights, and other factors that would make scene segmentation difficult. Objects are recognized by extracting 2-D features which are matched against 2-D object prototypes. This limits the system's recognition to known objects observed from standard viewpoints. Many tasks naturally fit these constraints or can be readily engineered to do so, with structured light and other artifices. Examples include picking parts off a moving conveyor, bonding leads to semiconductor chips, and inspecting bottles for misaligned labels.

In many industrial vision tasks the required engineering can be unacceptably expensive, difficult, or time consuming.

IV. Tasks Beyond Current Limitations

Bin Picking
Recognition of Parts on Overhead Conveyor
Recognition of Nonrigid Objects
Implicit Inspection
Robot Vehicles

Bin picking, for example, is hard because parts in a jumble have low contrast and occlude each other, making it difficult to isolate individual parts in an image. Inspecting parts on a finished assembly is difficult for the same reason and less amenable to engineering simplification (e.g., first dumping the contents of the bin on a table). Swinging parts on an overhead conveyor are not constrained to maintain a standard viewpoint. Nonrigid objects can assume a continuum of configurations and thus do not lend themselves to characterization in terms of fixed 2-D prototypes. Similarly, it is impractical to model in detail the appearance of all conceivable flaws (dents, scratches, blemishes and so forth) in an implicit inspection task. An archetypical example of a class of tasks that are inherently difficult to constrain are those involving robot vehicles in outdoor environments, such as construction site clearing and forresting.

Figures 1 - 4 provide more specific examples of important industrial vision requirements that are easy for humans to fulfill yet significantly beyond the current competence of machine vision. In Figure 1, the reader immediately recognizes the context, a bin-picking task, and is also cognizant of the layout of space; i.e., the area occupied by the bin and by the arm situated behind it. The representation of space facilitates complex spatial reasoning tasks such as planning approach-trajectories for the arm that avoid collisions with the sides of the bin.

Within the bin (Figure 2), the reader is aware of surfaces and the solid bodies they comprise. With such descriptions one can decide what part to grasp (based on shape and ease of access), how to pick it up (based on dimensions and inferred weight distribution), what the consequences of doing so will be (ie., what fragment of the scene is likely to lifted along with the grasped body and what other parts are likely to be disturbed), and how the selected part can be mated with others on a pallet or in an assembly. The contemplated grasping operation can then be performed using 3-D visual servoing to keep the manipulator on its trajectory and orient it properly (Figure 3). Note that none of these functions requires any special assumptions about lighting or viewpoint or any prior familiarity with the specific parts or arm involved.

It is, of course, possible to recognize bodies by matching them with known object models. Matching can occur either before or after a part has been removed from a bin, independent of viewpoint and occlusion; complex, articulated objects like the arm are recognized as readily as simple rigid ones. Common imperfections ranging from surface marring (rust and scratches) to missing limbs are recognized as defects (as distinguished from artifacts like shadows or errors in matching)--even though defects are not explicitly represented in the model (Figure 4).

Why are these tasks that are so easy for humans to perform currently so difficult for computers?

V. Competence Limited by:

Representations
Knowledge

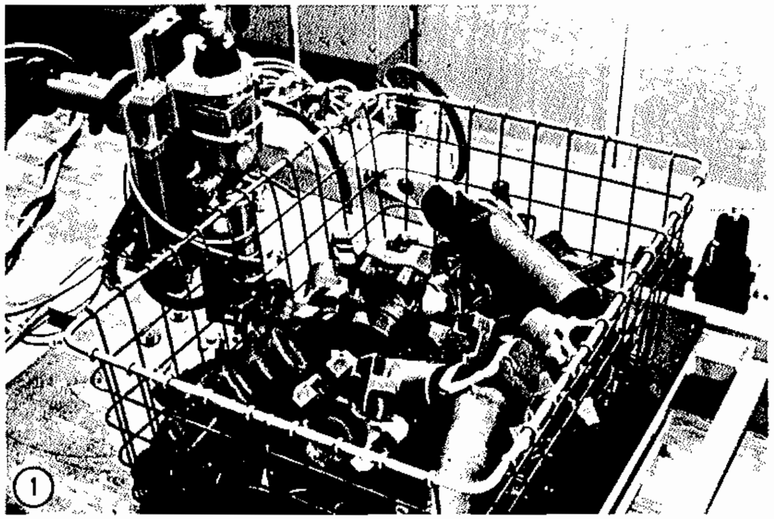
The competence of any vision system is limited by the representations it uses to describe the world and the knowledge available for manipulating and transforming them [1 and 2].

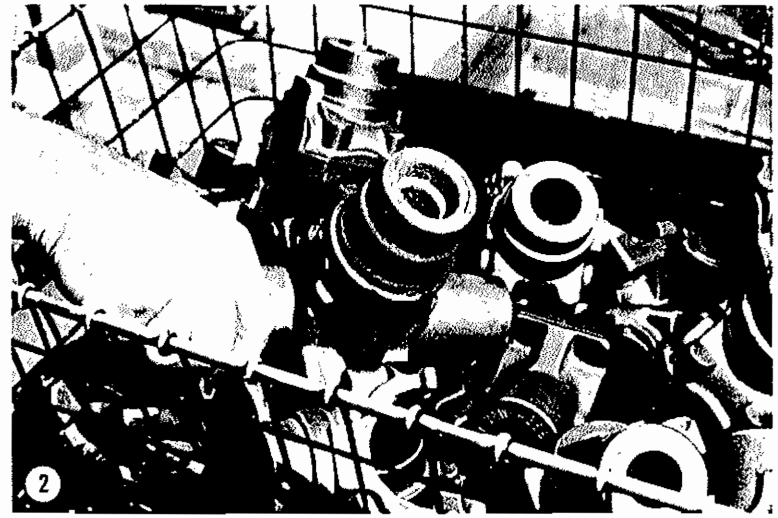
The specific capabilities illustrated in the bin-picking example establish minimum requirements for representation and knowledge.

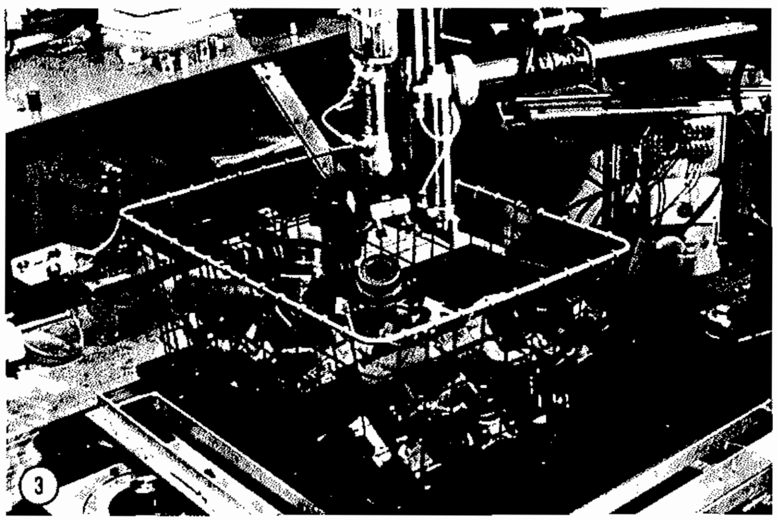
VI. Required Levels of Representation

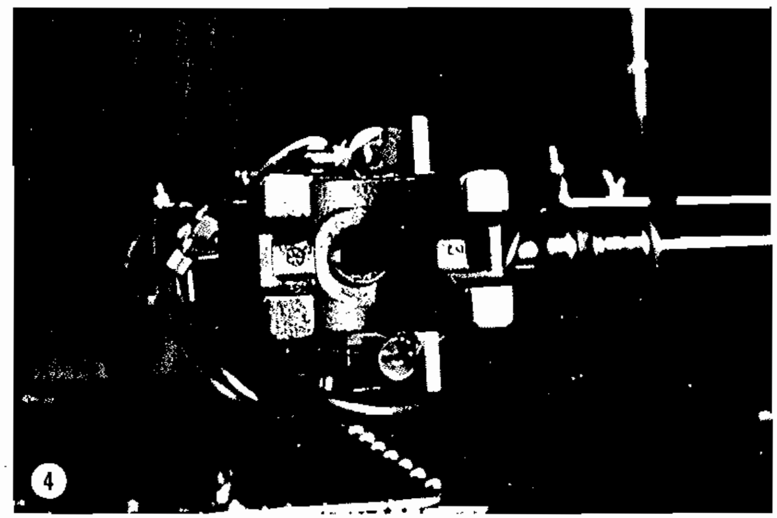
Images
Pictorial Features
Intrinsic Surface Characteristics
3-D Surfaces and Bodies
3-D Objects
Space Map
Symbolic Relationships

The ability to segment and describe the surfaces in a scene independent of lighting, shadows, and other viewing artifacts requires as a prerequisite the ability to describe the color, distance, orientation and other intrinsic characteristics of the surface element visible at each point in the image. The ability to inspect surfaces requires, in addition, a generic model of flaws: what is a dent? a scratch? a discoloration? The ability to grasp and manipulate parts without explicitly recognizing them demands a 3-D volumetric









FIGURES 1–4 BIN PICKING: A "TOUGH NUT" FOR INDUSTRIAL VISION SYSTEM

representation of bodies occupying a region of space, together with a generic understanding of what constitutes a manipulatable body, and how various physical forms should be grasped. The ability to recognize objects independent of viewpoint demands the ability to extract view-invariant 3-D features from the body descriptions and match them with features of 3-D object models. The ability to reach for and manipulate objects in a scene requires a representation of the 3-D layout of space and descriptions of mechanical relations between objects such as "Supports," "Touches" and "Interlocks."

The importance of most of these levels of description has been recognized for many years in scene analysis research. With one key exception, they were all explicitly represented in the hand-eye systems developed at Stanford and MIT circa 1969/70 [3].

By contrast, current industrial vision systems use very impoverished representations, relying heavily on detailed models of particular objects to accomplish tasks.

VII. Representations in Current Industrial Vision System

Images
Pictorial Features (Edges & Regions)
2-D Feature Attributes
Objects (Views)

Current systems often begin by thresholding the original grey-level image to obtain a binary array. Pictorial features (regions or edges) are extracted from the grey level or binary image and equated with surfaces and surface boundaries. Recognition is accomplished by matching 2-D attributes of these pseudo-surface features symbolically against 2-D models representing specific views of possible objects.

Given their impoverished representation, the gross inadequacies of current systems are hardly surprising.

VIII. Limitations of Current Systems

Weak Feature Extraction
No Invariance with Viewpoint
Restricted to Known Objects
No Descriptions of Surface Characteristics

Pictorial feature extraction is unreliable because region- and edge-finding programs have no basis for distinguishing which image features correspond to significant scene events (i.e., surfaces and surface boundaries) and which do not (i.e., shadows, highlights, etc.). Equally distressing are cases of low contrast where important surface boundaries are obscured by low contrast (see Figure 5).

Significantly, human perception of surface boundaries does not appear to depend critically on contrast. In Figure 6, for example, an intersection boundary is clearly perceived despite the absence of local contrast (Figure 7); its presence is demanded by the integrity of the surfaces it joins.

Subjective contour illusions, like the so-called sun illusion (Figure 8), appear to be an extreme example of this same phenomenon, where an edge is clearly perceived despite the complete absence of local evidence. A plausible explanation is that the edge corresponds to the boundary of an occluding disk-shaped surface, whose presence is implied by the abrupt line endings [4].

Even if an image could be reliably partitioned into surfaces based on regions and edges, the two dimensional shape features used for region description would still limit recognition to known objects observed from standard viewpoints; the ability to describe a new object so that it can be subsequently recognized from a different viewpoint requires three-dimensional description at the level of surfaces and volumes.

Current systems seldom have representations for interior characteristics of surfaces such as albedo, color, texture, and orientation. Such information is thrown away in thresholding to produce

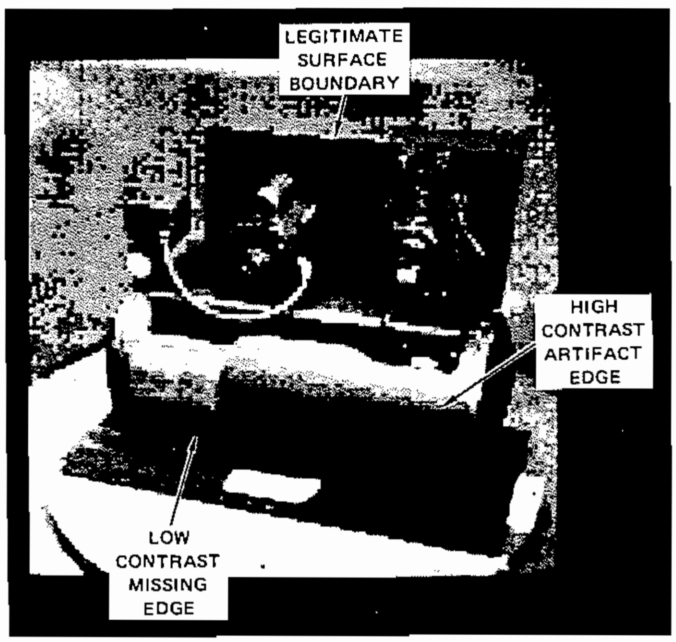


FIGURE 5 DIGITIZED IMAGE OF COMPRESSOR
(5 BITS AT 120 X 120 RESOLUTION)

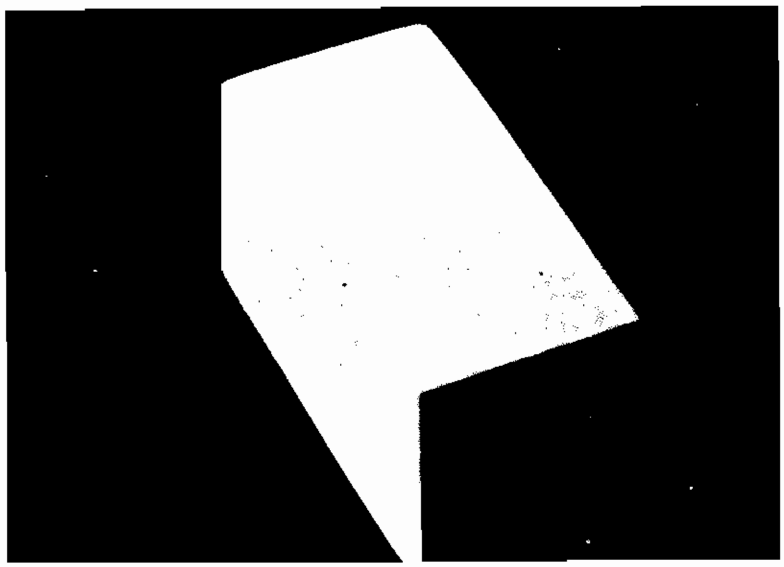


FIGURE 6 LOW CONTRAST INTERIOR BOUNDARY

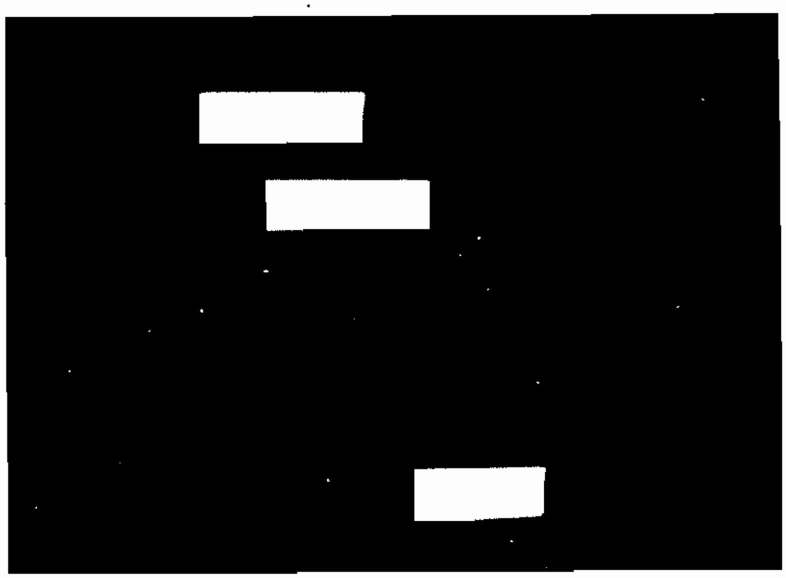


FIGURE 7 THE THREE HOMOGENEOUS STRIPS IN THIS FIGURE ARE TYPICAL CROSS SECTIONS OF THE SURFACE BOUNDARY IN FIGURE 6, VIEWED THROUGH A MASK. THERE IS VIRTUALLY NO LOCAL EVIDENCE FOR THE EDGE.

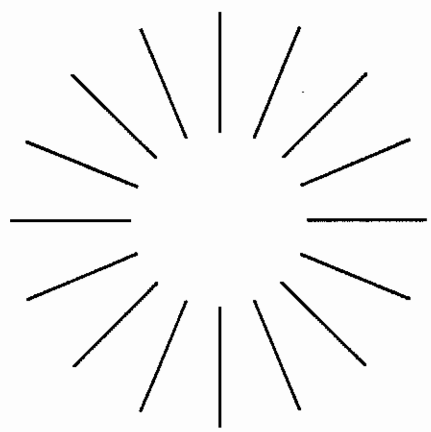


FIGURE 8   A SUBJECTIVE CONTOUR

a binary image. Even in systems that don't threshold, uniformity of grey level is usually equated with uniformity of surface appearance in inspection tasks. The folly of this oversimplification might be exemplified by an example such as a painting robot which doggedly attempts to achieve a uniform finish by repeatedly painting over its own shadow.

To summarize, the limitations of current systems stem directly from their attempts to deal with 3-D and physical characteristics of scenes in terms of 2-D pictorial features and 2-D object models. What is needed are many intermediate levels of representation to fill the enormous gap that exists between pictorial features and objects.

IX. Intrinsic Surface Characteristics

Images
Pictorial Features
*** Intrinsic Surface Characteristics ***
3-D Surfaces and Bodies
3-D Objects
Space Map
Symbolic Relationships

We would like to single out one level of description--intrinsic surface characteristics--which we believe is fundamental to many basic inspection and manipulation tasks and a prerequisite to obtaining the other missing levels of description. Specifically, we believe it is important to recover the range, orientation, reflectance, hue, and other physical characteristics of the three-dimensional surface element visible at each point in the image. These characteristics are intrinsic properties of the surface and are independent of lighting and other viewing artifacts.

Humans seem able to infer such information throughout a wide range of viewing conditions, even when the scene does not contain familiar objects and when deprived of such powerful cues as stereopsis, color, and motion. No current scene analysis system incorporates this level of description; we believe this is a fundamental reason why their performance is so inferior to that of humans.

Intrinsic characteristics are conveniently represented by a set of arrays in registration with the original sensed-image array. Each array contains values for a particular characteristic of the surface element visible at the corresponding point in the sensed image. Each array also contains explicit indications of boundaries due to discontinuities in value or gradient. We call such arrays Intrinsic Images.

The primary intrinsic images are of surface reflectance, surface orientation, and incident illumination. Other characteristics, such as range, transparency, specularity, and so forth, might also be useful as intrinsic images, either in their own right or as intermediate results. The distance and orientation images together correspond to Marr's notion of a 2.5D sketch [2].

Figure 9 gives an example of one possible set of intrinsic images corresponding to a monochrome image of a simple scene.

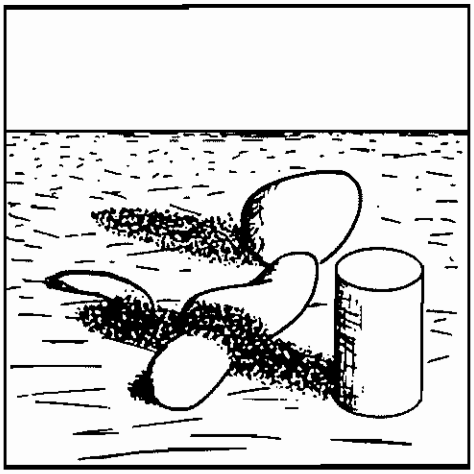
A concrete example of intrinsic images and their usefulness in computer vision can be seen in experiments by Nitzan, Brain, and Duda with a scanning laser range finder which directly measures the intrinsic properties of distance and apparent reflectance.

Because the range data is uncorrupted by reflectance variations and the amplitude data is unaffected by ambient lighting and shadows, it is easy to extract surfaces of uniform height (Figure 10c) or reflectivity (Figure 10e) and surface boundaries where range is discontinuous (Figure 10d). Such tasks are very difficult to perform reliably in grey-level imagery; but with pure range and amplitude data, even simple-minded techniques such as thresholding and region-growing work well.

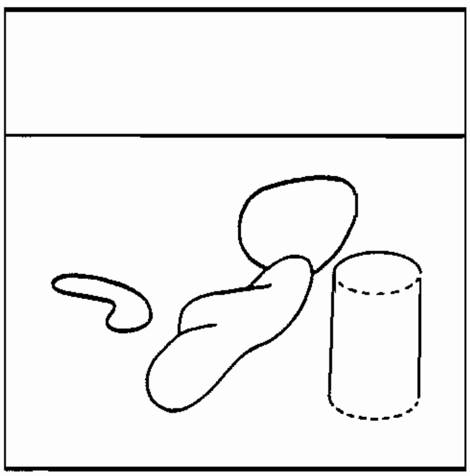
Laser range finders may eventually make good industrial sensors; but they are currently expensive, slow, and not appropriate in all applications. It is thus important to understand how it may be possible to recover intrinsic characteristics from ordinary grey-level imagery, as humans apparently do.

FIGURE 9 A SET OF INTRINSIC IMAGES DERIVED FROM A SINGLE MONOCHROME INTENSITY IMAGE

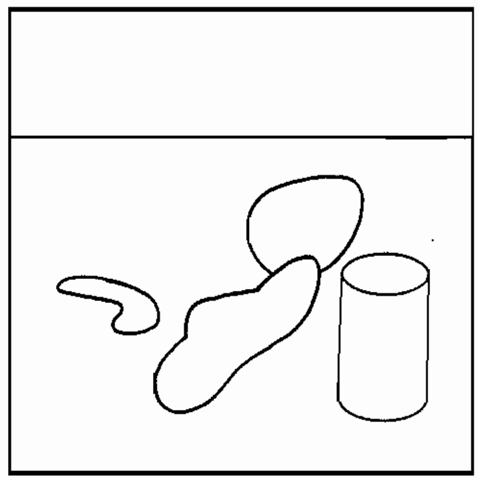
The images are depicted as line drawings, but, in fact, would contain values at every point. The solid lines in the intrinsic images represent discontinuities in the scene characteric; the dashed lines represent discontinuities in its derivative. In the input image, intensities correspond to the reflected light flux received from the visible points in the scene. The distance image gives the range along the line of sight from the center of projection to each visible point in the scene. The orientation image gives a vector representing the direction of the surface normal at each point. The reflectance image gives the albedo (the ratio of total reflected to total incident illumination) at each point.



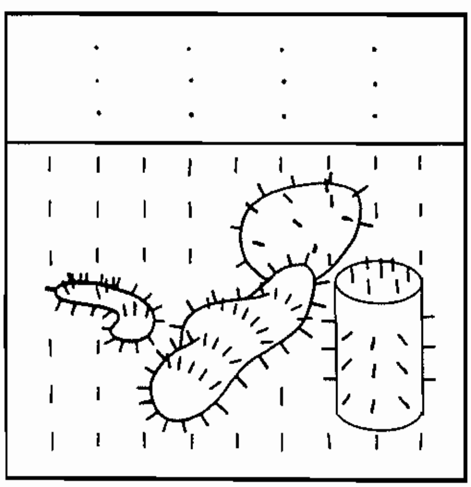
(a) ORIGINAL SCENE



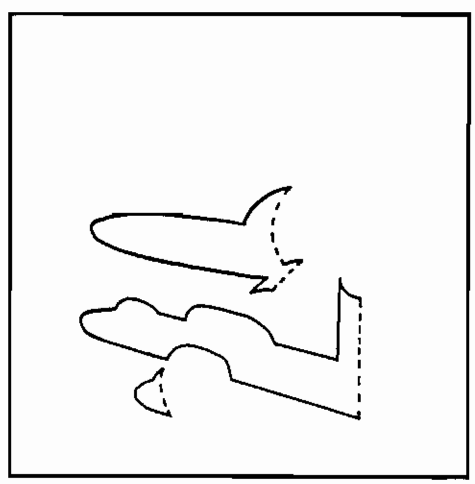
(b) DISTANCE



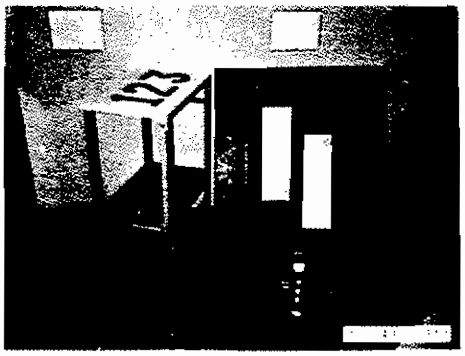
(c) REFLECTANCE



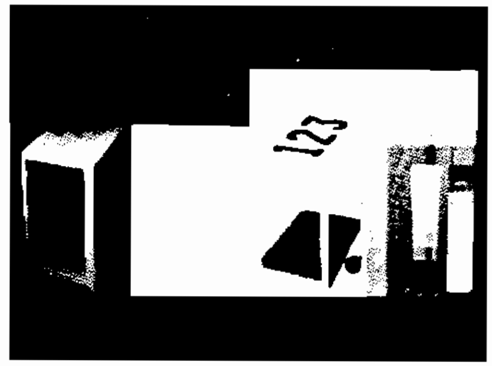
(d) ORIENTATION (VECTOR)



(e) ILLUMINATION

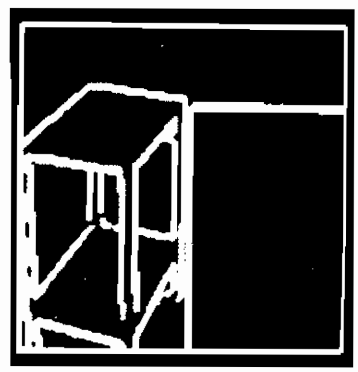


(a) A CONVENTIONAL PHOTO OF A SCENE

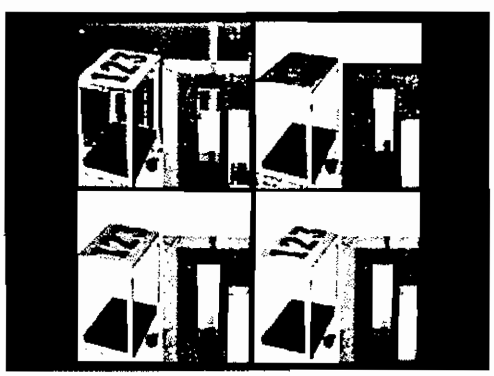


(b) DISTANCE AND REFLECTANCE IMAGES

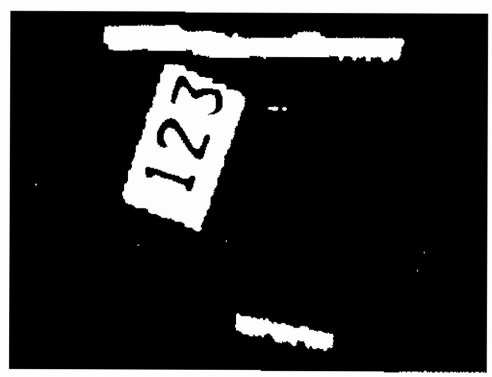
(c) EXTRACTED PLANAR SURFACES



(d) DISCONTINUITIES IN RANGE



(e) THRESHOLDING REFLECTANCE



(f) CORRECTED VIEW OF CART TOP

FIGURE 10 EXPERIMENTS WITH A LASER RANGE FINDER

X. Recovery of Intrinsic Characteristics

The central problem in recovery is that the desired information is confounded in the sensory data. When an image is formed, whether by a camera or an eye, the light intensity at each point in the image is determined by the incident illumination, the surface reflectance, the surface orientation, and other intrinsic characteristics of the corresponding point in the scene. In the simple case of an ideally diffusing surface, for example, the image light intensity, L, is given by Lambert's Law:

$$L = I * R * \cos i$$

where I is intensity of incident illumination, R is reflectivity of the surface, and i is the angle of incidence of the illumination (see Figure 11).

Since the observed light intensity at a single point could result from any of an infinitude of combinations of illumination, reflectance and orientation, one might be tempted to conclude that unambiguous recovery is impossible. However, recent research at SRI, building on work by Marr and Horn at MIT, has led us to believe that the information can be unscrambled, at least in simple domains like that depicted in Figure 9.

The secret of recovery necessarily lies in exploiting constraints derived from assumptions about the nature of the scene and the physics of the imaging process. Since surfaces are continuous in space, their characteristics (reflectance, orientation, range) are generally continuous across an image, except at surface boundaries. Incident illumination also usually varies smoothly over a scene, except at shadow boundaries. Elementary photometry tells us that where all intrinsic characteristics are continuous, image brightness is continuous; conversely, where one or more intrinsic characteristics are discontinuous, a brightness edge will usually result.

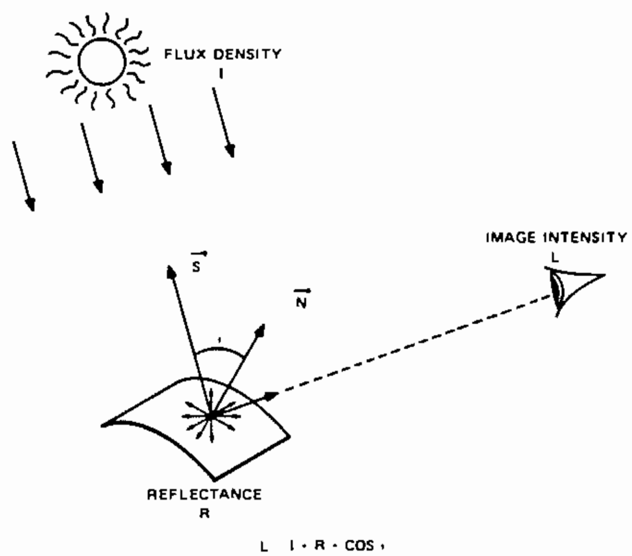


FIGURE 11 AN IDEALLY DIFFUSING SURFACE

The pattern of brightness variation in an image can provide important clues to the local behavior of the intrinsic characteristics. A well-known example is the determination of surface shape from shading for surfaces of uniform reflectance under distant point-source illumination [5].

In [5] it was assumed that image intensity was continuous and that variations were due solely to variations in surface orientation. When a scene contains multiple objects, that may be differently colored, or cast shadows, such assumptions are invalid. One is then faced with the problem of deciding what physical characteristic (or characteristics) is, in fact, responsible for an observed intensity variation and which characteristics are discontinuous across intensity edges.

The pattern of brightness variation on either side of an intensity edge can sometimes provide strong clues to the type of scene event responsible, (shadow or surface boundary) and thus to which intrinsic characteristics are actually discontinuous at that point. A simple example is the characteristic penumbrae and high contrast of shadow edges, indicating a discontinuity in illumination. The interpretation of brightness edges as scene events is also important because knowing the type of scene event sometimes allows explicit values to be determined for some of the intrinsic characteristics. For example, at an extremal occluding boundary, where an object curves smoothly away from the viewer, the surface orientation can be inferred exactly at every point along the boundary. A test for extremal boundaries can be made by determining whether the observed brightness variation along the edge is consistent with the expected orientation.

We have compiled an exhaustive catalog of edge interpretations for the simple scene domain of Figure 9, together with the constraints and values implied by each interpretation (Table 1).

The ability to classify edges in this way, at least for simple domains, suggests the following recovery paradigm:

Table 1 THE NATURE OF EDGES

LA and LB refer to variations of intensity along sides A and B of an edge. Intensities are either constant, varying, or varying in accordance with the assumed orientations along an extremal boundary, the so-called tangency condition.

Region Intensities LA	 LB	Edgs Type	Region Types	Intrinsic Edges Intrinsic Values D	 N	 R	 I
Constant	Constant	Occluding sense unknown	A B shadowed	EDGE	EDGE	EDGE RA RB	IA IB
Constant	Varying	1 Shadow	A shadowed B illuminated		NB.S	RA RB	EDGE IA IB
		2 A occludes B	A shadowed B illuminated	EDGE DA<DB	EDGE NA	EDGE RA	EDGE IA
Varving	Varying	Inconsistent with domain					
Constant	Tangency	B occludes A	A shadowed B illuminated	EDGE DA>DB	EDGE NB	EDGE RA RB	EDGE IA IB
Varying	Tangency	B occludes A	A B illuminated	EDGE DA>DB	EDGE NB	EDGE RB	EDGE IB IA
Tangency	Tangency	Not seen from general position					

XI. Steps in Recovery

Edge Detection
Edge Classification
Establishment of Boundary Conditions
Propagation of Boundary Values into Regions
(Edge Refinement)

Edges are extracted in the intensity image and classified according to the catalog. The edge interpretations may provide values for one or more intrinsic characteristics at surface points along one or both sides of the edge. Values at interior surface points are determined by propagating in these boundary values, obeying continuity assumptions. Discontinuities may occur where values propagated in from opposite boundaries meet. This most often occurs when a physical edge does not result in a visible intensity discontinuity (as was the case in Figure 6). In such events, missing edges can be hypothesized, establishing new boundary conditions, and the propagation process repeated, until a consistent set of boundaries and values are obtained.

The ability to refine the initial edge description as an integral part of the recovery process is essential to a practical theory. Extracting an ideal line drawing in a grey-level image is known to be very difficult, if not impossible. However, it is usually possible to extract a fairly good approximation containing a few gaps or extraneous lines. All that is required by the theory is that the initial line drawing be good enough so that any errors will show up as inconsistencies during the propagation process.

Another attractive feature of the model is its dependence on local computations which can be performed rapidly in parallel by low-cost LSI processor arrays. One possible implementation is sketched in Figure 12. In essence, it consists of a stack of registered arrays representing the original intensity image (top) and the primary intrinsic arrays. Processing is initialized by detecting intensity edges in the original image, interpreting them according to a catalog of appearances, and then creating the appropriate edges in the intrinsic images (as implied by the descending arrows).

Parallel local operations (shown as circles) modify the values in each intrinsic image to make them consistent with intra-image continuity and limit constraints (for example, reflectance must be between 0 and 1). Simultaneously, a second set of processes (shown as vertical lines) operates to make the values at each point consistent with the corresponding intensity value, as required by the inter-image photometric constraint. A third set of processes (shown as Xs) operates to insert and delete edge elements, which inhibit continuity constraints locally. The constraint and edge modification processes operate continuously and interact to recover accurate intrinsic scene characteristics and to perfect the initial edge interpretation.

The action is envisaged to resemble an analog computer: as the value in one image increases, corresponding values in other images increase or decrease to maintain consistency with the observed intensity at that point. Within each image, values tend to propagate in from boundary conditions established along edges. This resembles relaxation processes used in physics for determining temperature or potential over a region from boundary conditions.

Our model can be regarded as a generalization of Horn's lightness model [6], Woodham's shading model [8] and Marr and Poggio's cooperative stereopsis model [7] for simultaneous recovery of geometric and photometric attributes. The model is significant, despite the simplicity of the domain, because it demonstrates for the first time the theoretical possibility of simultaneously recovering orientation, reflectance, and illumination from a single monochrome image, without recourse either to object prototypes or to primary depth cues, such as stereopsis, motion parallax, or texture gradient. We believe that these and many other well-known sources of information could be incorporated to improve overall performance, consequently extending the model's competence to deal with more complex scenes.

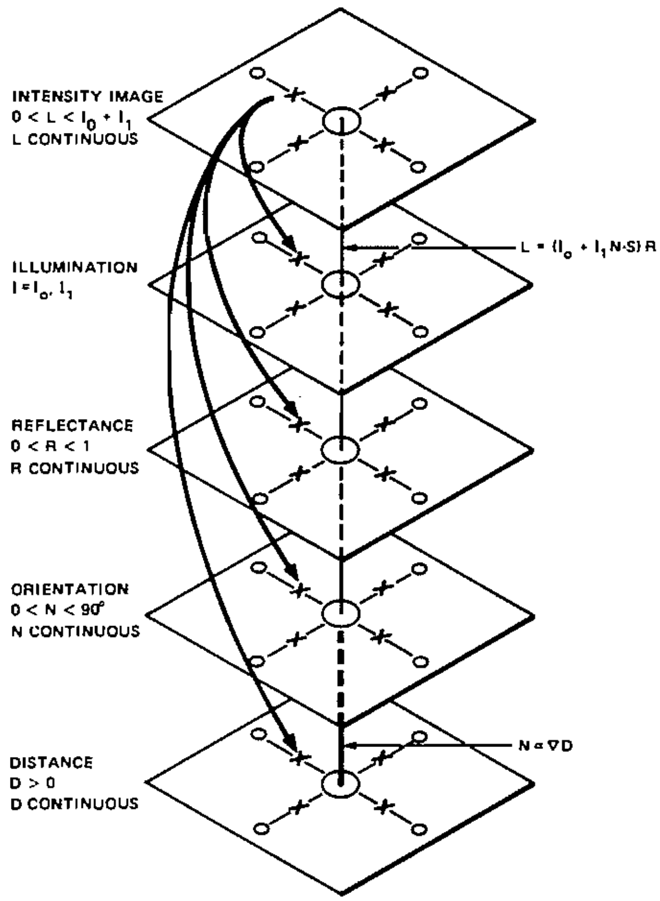


FIGURE 12 A PARALLEL COMPUTATIONAL MODEL FOR RECOVERING INTRINSIC IMAGES

In particular, the current model can easily capitalize on simplifications that are often available in an industrial context. For example, it may be possible, by controlling lighting, to initialize values in the incident illumination array. Similarly, the reflectance array can be initialized when the albedo of a surface is known. The orientation array can be initialized using active techniques such as grid coding [9] or photometric stereo [10], and range can be determined with a laser range finder.

XII. Exploiting Intrinsic Characteristics

Detection of Local Anomalies
Interpretation of Anomalies as Scratches, Dents, Holes ...
Segmentation into Surfaces and Bodies
Extraction of Invariant 3-D Shape Features
Object Recognition Based on 3-D or Generic Models
Material Recognition Based on Reflectance
Spatial Reasoning

Returning to the main theme of this paper, the key idea is that intrinsic images, no matter how recovered, provide direct access to 3-D and photometric properties of a scene that are not commonly available in computer vision systems. Such information facilitates all subsequent levels of description (eg. surfaces, volumes) and opens up exciting new areas for research.

At SRI, for example, we have an active project concerned with segmenting range and reflectance arrays from the range finder into surfaces and volumes. Surprisingly, finding surfaces is not quite as trivial as one might expect; range is often continuous across surface boundaries (eg., where floors meet walls) so that surface boundaries must be detected by discontinuities in local orientation. Since orientation is related to the derivative of range, it can be quite noisy (in practice a 1% range uncertainty can result in up to 30 degrees of orientation uncertainty). Thus orientation must be smoothed by spatial averaging. However, one must be careful not to average over the surface boundaries one is attempting to find. Hence, smoothing and boundary finding present a chicken and egg situation.

The above difficulties suggest that volumes, rather than surfaces, may be better candidates for initial scene partitioning because small perturbations in range have little effect on shape, symmetry axes, and other global characteristics of volumes.

Volumes are defined by surface elements that surround compact regions of space. Often these elements cluster about a symmetry axis. Accordingly, we are thinking about using a 3-D symmetric axis transform to locate probable axes from the range finder data and representing the associated volumes with generalized cylinders [11]. Such a volume representation could then be matched to 3-D object models, building on the work of Nevatia and Binford [12] and of Marr and Nishihara [13].

The immediate possibility of recovering intrinsic images by mechanical or other means suggests many other interesting projects. These include detecting local anomalies and interpreting them generically as scratches, dents, discolorations and so forth, recognizing objects from generic descriptions (i.e., the type of description that allows recognition of all manner of nuts in Figure 13), recognizing materials (wood, plastic, metal), and planning manipulation trajectories using the space map.

XIII. Conclusions

Although many industrial tasks fall within the competence constraints of current machine-vision systems, many others require a more general-purpose, high-performance vision capability approaching that of humans. In our current view, such a system would be based on an initial level of processing (Figure 14) that transforms the information in the sensed image into an intermediate level of representation we call intrinsic images. Intrinsic images provide access to the physical and three-dimensional properties of a scene and thereby greatly facilitate all subsequent levels of perceptual processing.

In the near term, industrial vision systems can rely on active sensors that directly measure intrinsic characteristics. Therefore,

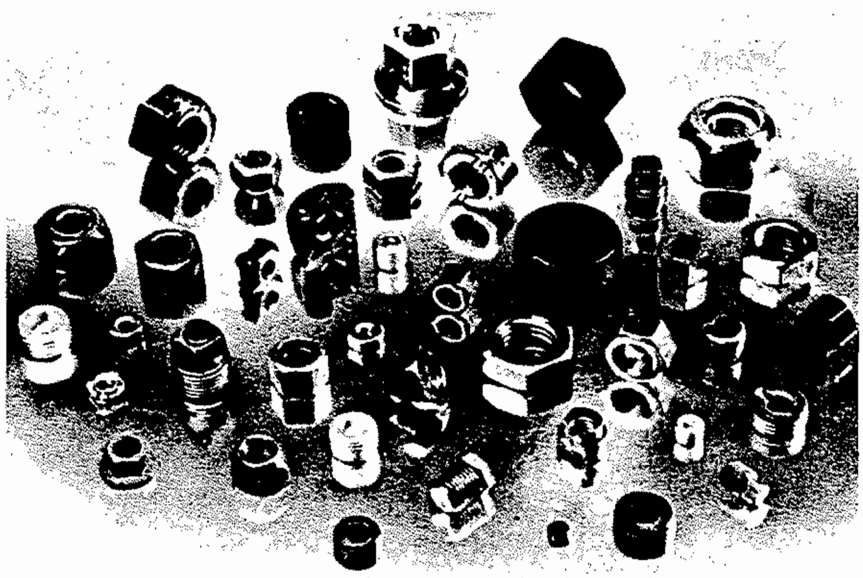


FIGURE 13 MOST READERS WILL HAVE LITTLE DIFFICULTY RECOGNIZING THE NUTS IN THIS IMAGE; DESPITE THE VARIETY OF SHAPES, SIZES AND COLORS.

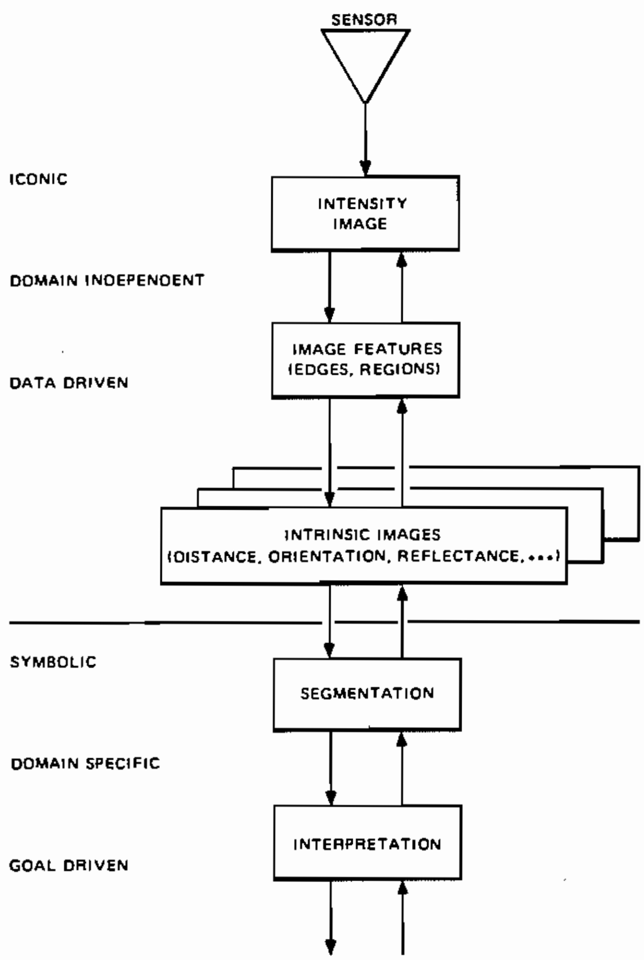


FIGURE 14 ORGANIZATION OF A VISUAL SYSTEM

researchers should now be seriously considering how to exploit this level of description. Ultimately, we hope to show the feasibility of recovery from ordinary grey-level video imagery, which would make many more applications practical.

XIV. Key Ideas

Multiple Levels of Representation
Intrinsic Surface Characteristics
3-D Description and Object Models
Generic Models

Other key ideas in this paper include the need for multiple levels of representation that make explicit the information that must be recovered for each task, the use of 3-D models to achieve independence of viewpoint, and the use of generic models to allow manipulation, inspection, and description without explicit recognition. A system with these capabilities would exhibit much of the flexibility and robustness that makes human vision invaluable in factories today.

REFERENCES

1. H. G. Barrow and J. M. Tenenbaum, "Representation and Use of Knowledge in Vision," Technical Note 108, Stanford Research Institute, Menlo Park, California (July 1975); an expanded version of a paper that originally appeared in Sigart, No. 52 (June 1975).

2. D. Marr, "Representing Visual Information," in Computer Vision Systems, A. Hanson and E. Riseman, eds. (Academic Press, New York, New York, 1978).

3. J. A. Feldman et al., "The Stanford Hand-Eye Project," Proc. IJCAI, pp. 521-526 (May 1969).

4. H. G. Barrow and J. M. Tenenbaum, "Recovering Intrinsic Scene Characteristics from Images," Technical Note 157, SRI International, Menlo Park, California (April 1978); to appear in Computer Vision Systems, A. Hanson and E. Riseman, eds. (Academic Press, New York, New York, in press).

5. B.K.P. Horn, "Obtaining Shape from Shading Information," in The Psychology of Computer Vision, P. H. Winston, ed. (McGraw-Hill, New York, New York, 1975).

6. B.K.P. Horn, "Determining Lightness from an Image," Computer Graphics and Image Processing, Vol. 3, pp. 277-299 (1974).

7. D. Marr and T. Poggio, "Cooperative Computation of Stereo Disparity," Science, Vol. 194, pp. 283-287 (1977).

8. R. J. Woodham, "A Cooperative Algorithm for Determining Surface Orientation from a Single View," Proc. Fifth Intl. Joint Conference on Artificial Intelligence, Cambridge, Massachusetts, pp. 635-641 (August 1977).

9. P. M. Will and K. S. Pennington, "Grid Coding: A Preprocessing Technique for Robot and Machine Vision," Second Intl. Joint Conference on Artificial Intelligence, Imperial College, London, England, pp. 66-70 (September 1971).

10. Fabio Metelli, "The Perception of Transparency," Scientific American (April 1974).

11. G. J. Agin and T. O. Binford, "Computer Description of Curved Objects," Proc. Third International Joint Conference on Artificial Intelligence, Stanford University, Stanford, California, pp. 629-640 (August 1973).

12. R. Nevatia and T. O. Binford, "Perception and Recognition of Curved Objects," Artificial Intelligence, Vol. 8 (1977).

13. D. Marr and H. K. Nishihara, "Representation and Recognition of the Spatial Organization of 3-D Shapes," Proc. Roy. Soc. Lond., Vol. 200, pp. 269-294.