ornl

OAK RIDGE NATIONAL LABORATORY

MARTIN MARIETTA

ORNL/TM-11349
(CESAR-89/38)

Treatment of Systematic Errors II: Fusion of Ultrasound and Visual Sensor Data

M. Beckerman
L. A. Farkas
S. E. Johnston

OPERATED BY
MARTIN MARIETTA ENERGY SYSTEMS, INC.
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

ORNL/TM-11349
CESAR-89/38

Engineering Physics and Mathematics Division

TREATMENT OF SYSTEMATIC ERRORS II: FUSION OF ULTRASOUND AND VISUAL SENSOR DATA

M. Beckerman, L. A. Farkas and S. E. Johnston

Research sponsored by the Engineering Research Program of
the Office of Basic Energy Sciences, and the Office of
Nuclear Energy, Office of Technology Support Programs
of the U.S. Department of Energy

Published: June 1990

Prepared by the
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
operated by
MARTIN MARIETTA ENERGY SYSTEMS, INC.
for the
U.S. DEPARTMENT OF ENERGY
under contract DE-AC05-84OR21400

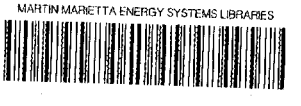

TABLE OF CONTENTS

INDEX TERMS

Autonomous Robots

Mobile Robots

Object Recognition

Sensor Fusion

Systematic Errors

Ultrasound Sensing

Visual Sensing

World Modeling

ABSTRACT

In this work we present a methodology for the fusion of ultrasound and visual sensor data as acquired by a mobile robot. The objective of the methodology was the reduction of systematic errors which arise in the processing of the data in the individual sensor domains. In the initial processing of the ultrasound scan, rectilinear (Cartesian map) and polar (strings) data structures were built. In the initial processing of the CCD camera image, vertical edge segments were identified and labelled according to their connectivity. The systematic errors treated included ultrasound distortions in size, and visual ambiguities in discriminating depth discontinuities from intensity gradients generated by other details in the image. These systematic errors were first flagged by comparing the ultrasound strings and visual vertical edges to one another. The ranges, spatial orientation of the camera, and geometric information extracted from least-squares fits were then used in the fusion stage processing of the visual image. Vertical edge information was used in the subsequent fusion stage processing of the ultrasound data. By the end of this feedback-like fusion process the data structures in each sensor domain carried some information from the other domain. We had identified the vertical edges of interest, tagged them with range information, and removed the distortions from the Cartesian navigation maps.

1. INTRODUCTION

One of the prerequisites for intelligent behavior in robotic systems is the ability to generate internally self-consistent representations of the environment from sensor data. Often, this is impossible to achieve on the basis of either one sensing operations or a single sensor domain, particularly when the environment is unstructured to any appreciable extent. The term "multi-sensor integration", or equivalently, "multi-sensor fusion", denotes the task of combining data and information from more than one robot location, and/or from more than one sensor, so that a consistent world model can be generated.

The term "systematic errors" encompasses all errors which are not random in character (i.e., which are not due to, say, electronic noise or counting statistics, etc.). These errors are produced whenever there are incorrect and/or inconsistent interpretations of data. They arise in automated, or robotic, systems when there is insufficient sensor information to correctly interpret the data, and when assumptions concerning the nature of the environment are violated.

Multi-sensor fusion can be viewed as a strategy for the reduction of systematic errors to produce a self-consistent internal representation of the environment. There are two logical stages in this type of strategy. In the first phase, consistency checks are done between different sets of sensor data which have been acquired and initially interpreted. In the second part, models of the sensors, the environment, and their mutual interactions, are used in order to resolve the inconsistencies found in the first stage of the fusion process. An example of this approach was presented in [4]. In that work ultrasound data from different robot locations were fused to produce Cartesian navigation maps. The cells of the Cartesian maps were assigned labels denoting whether they were empty or occupied. During the updating procedure, comparisons were made between the sets of labels assigned from data at different sensing locations. Inconsistencies were flagged by conflicting label assignments. Pattern analyses and consistent-labelling algorithms, based upon physical arguments and models, were then used to resolve the conflicting label assignments.

Turning to the growing literature on sensor fusion we note that some of the problems associated with achieving location self-consistency by mobile robots were discussed by Chatila and Laumond [7], and by McKendall and Mintz [21]. We observe that Ayache and Fauguras [3] investigated problems associated with building and fusing 3-D noisy visual maps, and Elfes [11] constructed 2-D maps from ultrasound data. Striped light and (passive) intensity images were integrated by Wang and Aggarwal [30] and by Hu and Stockman [16]; registered range and reflectance data from a laser rangefinder were fused by Duda, et al. [10], Nitzan et al. [25], and Magee et al. [19]. In addition, Flynn [12] studied the integration of ultrasound and infrared range sensor data, Allen [1], and Stanfield [29], investigated the integration of vision and touch, and Nandhakumar and Aggarwal [23] examined the fusion of thermal and visual sensor data.

In the present work we developed a methodology for fusing 2-D ultrasound range data with 2-D visual intensity data as acquired by a mobile robot. As was the case in [4], the goal of the fusion process was reduction of systematic errors. Although the 2-D planes in the ultrasound and visual sensor domains were different, the former referring to a horizontal scanning plane and the latter denoting a vertical image plane, the two sensor domains shared a common angular coordinate. This overlap

enabled us to use the ultrasound information reduce systematic errors in the visual domain, and vice versa.

The information we wished to extract from the data consisted of the positions in the visual image of the vertical edges of objects, and the locations in the ultrasound viewing plane of the occupied and empty regions of space, and also the angular positions of the depth discontinuities. These particular features were chosen for their utility with respect to robotic exploration and navigation and for their comparability. The processing of the data in the individual sensor domains to extract this information is described in Section 3. This discussion is preceded in Section 2 by a brief overview of the sensor environment and systematic errors.

We begin Section 4 by noting that our comparison-correction mechanism for fusion is analogous to a feedback process. In the case of sensor fusion we are attempting to control the interpretations of the data. The two feedback-like loops used in the fusion of ultrasound and visual sensor data are introduced in Section 4.1. The sequential execution of these loops is described in Sections 4.2 and 4.4, and Section 4.3 contains a discussion of the statistical analyses and spatial filtering done during the first part of the fusion process. The paper is summarized in Section 5.

2. SENSING, ENVIRONMENT AND SYSTEMATIC ERRORS

2.1. Sensor Placement and Data Acquired

Hermies-IIB is one of a series of mobile robot prototypes developed at Oak Ridge National Laboratory's Center for Engineering Systems Advanced Research (CESAR). The hardware for this robot includes a mobile platform, dual manipulator arms, and a sensor turret equipped with an array of twenty-four ultrasonic range sensors and several CCD cameras. This robot provided the data for the fusion study.

The ultrasonic sensors contained Polaroid industrial-grade transducers producing 50 kHz bursts, 1 msec in duration. The transducers functioned as both transmitters and receivers of ultrasound. By processing the signals to give the time of flight for the earliest echo the distance to the nearest object lying either partially or wholly within the beam cone was determined.

The twenty-four ultrasonic sensors were grouped into six phased-arrays of four units each. One phased-array was oriented in the forward direction for reflexive collision avoidance. The five remaining phased-arrays were mounted on the sensor turret to form a partial ring in the horizontal plane. By panning the sensor turret in one of several modes, range data were acquired spanning 360 deg. In our studies from 24 to 120 samples of range data were taken in 15 to 3 deg steps, respectively.

The intensity pattern of the phased-array configuration possesses a minimum (null) at 4.8 deg [18]. Due to mechanical limitations of the sensor, cancellations were incomplete and there were small side lobes in the intensity pattern at larger angles. The effective beam width of the sensors depends upon the radiated power, the surface properties of the surfaces scanned, and the amplifier threshold sensitivity. In the experiments performed by us the effective beam width of the sensors was approximately 18 deg.

The visual sensor used in the study consisted of a single Sony CCD camera equipped with a wide-angle, 4.8mm lens. This camera provided 256 x 256 pixels of 8-bit grey-level intensity values from a 60 degree field-of-view. The camera was mounted on the robot at a height of 94 cm above the floor, and was oriented facing forward with its optical axis in the horizontal plane.

2.2. Sensing Environment and Systematic Errors

Displayed in Fig. 1 is a visual scene depicting part of the CESAR laboratory which served as our experimental testbed. A single box has been placed in arbitrary position on the floor away from walls. The horizontal dimensions of the sides of this and other boxes used in the experiments varied from 30 to 120 cm, and were of the same order of magnitude as the ultrasound beam width at typical scanning distances. The boxes were tall enough to intercept the Hermies-IIB ultrasound viewing (horizontal) plane.

We observe in Fig. 1 that the floor has been marked with a rectangular grid. There is also a variety of wall markings and other background clutter. The upper portion of the figure shows the placement of unmodified, hanging fluorescent lamps which serve as a non-uniform source of ambient light for the CCD camera.

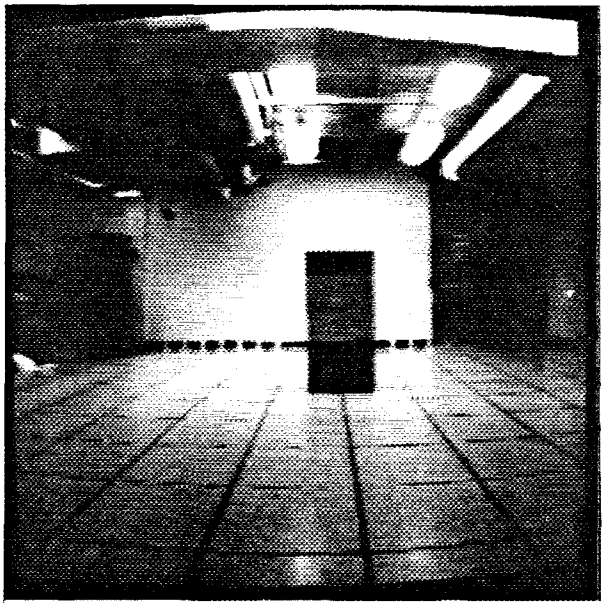

Fig. 1. Experimental setup: CCD image of part of the CESAR laboratory. The box to be identified is located in center of the image.

The ultrasonic systematic errors depend upon the beam width (resolution), the sensing frequency, the radiated power and sensing thresholds, and the environmental geometry and surface properties of the objects being scanned. There are two broad classes of errors. These are (i) distortions, and (ii) specular reflections/complete absorption. The former are resolution, power and threshold dependent; the latter are environmental geometry, surface property and frequency related. We focussed our attention in this study on the distortion errors; the second class of errors will be discussed by us elsewhere.

Several techniques for extracting edges and lines from CCD camera images have been developed. Roberts [27], Rosenfeld and Thurston [28], Nevatia and Babu [24], Marr and Hildreth [20], and Canny [6] for example, employed either 1st derivative operators or Laplacians; Marr and Hildreth [20], Yuille and Poggio [32], and Haralick [15] used zero crossings in the 2nd directional derivative, and Prewitt [26], Hueckel [17], Haralick [15] and Nalwa and Binford [22] developed surface-fitting methods.

Shown in Fig. 2 are the results of using a first derivative operator to find the vertical edges in the grey-scale CCD image. Both first and second derivative, directional and non- directional, edge detectors function by finding intensity gradients in the image (see, for example, the studies by Canny [6], Marr and Hildreth [20], Nalwa and Binford [22], and Haralick [15]). As a result, there will be systematic errors, arising from the interpretation of the output from these gradient operators as object boundaries, that is, as depth discontinuities. For example, we observe in Fig. 2 that for our environment we have depth discontinuities which characterize the boxes, and

we also have intensity gradients generated by variations in the background and floor markings, by the overhead lights and air- conditioning ducts, and by shadows and glare. The number of vertical edges in the processed image, in fact, far exceeds the number of actual depth discontinuities forming boundaries of the physically interesting objects.

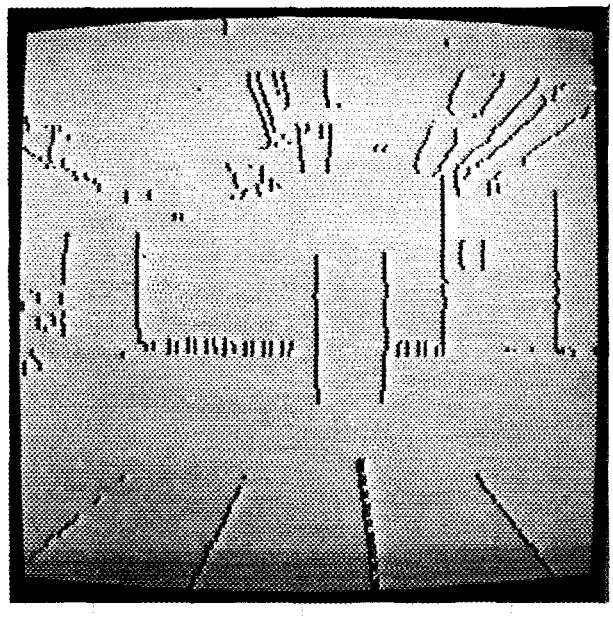

Fig. 2. Vertical edge segments. Features were extracted using anisotropic Gaussian smoothing and a horizontal derivative operator.

In the psychophysical theory of monocular vision [13,14], scenes such as that shown in Fig. 1 contains sufficient information for a spatial understanding of its features. That is, the textural, shading, perspective and edge information contained in the visual scene provides the (human) viewer with sufficient depth information to disambiguate the objects of interest. The minimally processed image displayed in Fig. 2 no longer contains this wealth of information. One solution to this lack of information would be to perform a more detailed processing of the initial data in order to extract textural, shading and other types of information. Alternatively, we observe that the minimal, vertical edge information is sufficient for navigation and exploration. Furthermore, we can provide the missing depth perception by making use of the information available in the ultrasound sensor domain, that is, by fusing the ultrasound and visual sensor data.

3. PROCESSING OF THE SENSOR DATA PRIOR TO FUSION

3.1. Initial Processing of the Ultrasound Scan

The initial processing of the ultrasound scan data is depicted in the block diagram, Fig. 3. In brief, in the pre-processing stage, consistency checks were done between returns from adjacent scan angles, and synthetic beams were generated at intermediate scan angles. The synthetic beams enabled the ray-tracing algorithm used in the mapping stage to cover the appropriate region of the Cartesian map. In the present work the initial sample of range data was transformed at the end of the pre-processing stage into 240 samples of range data spanning 360 deg in 1.5 deg steps.

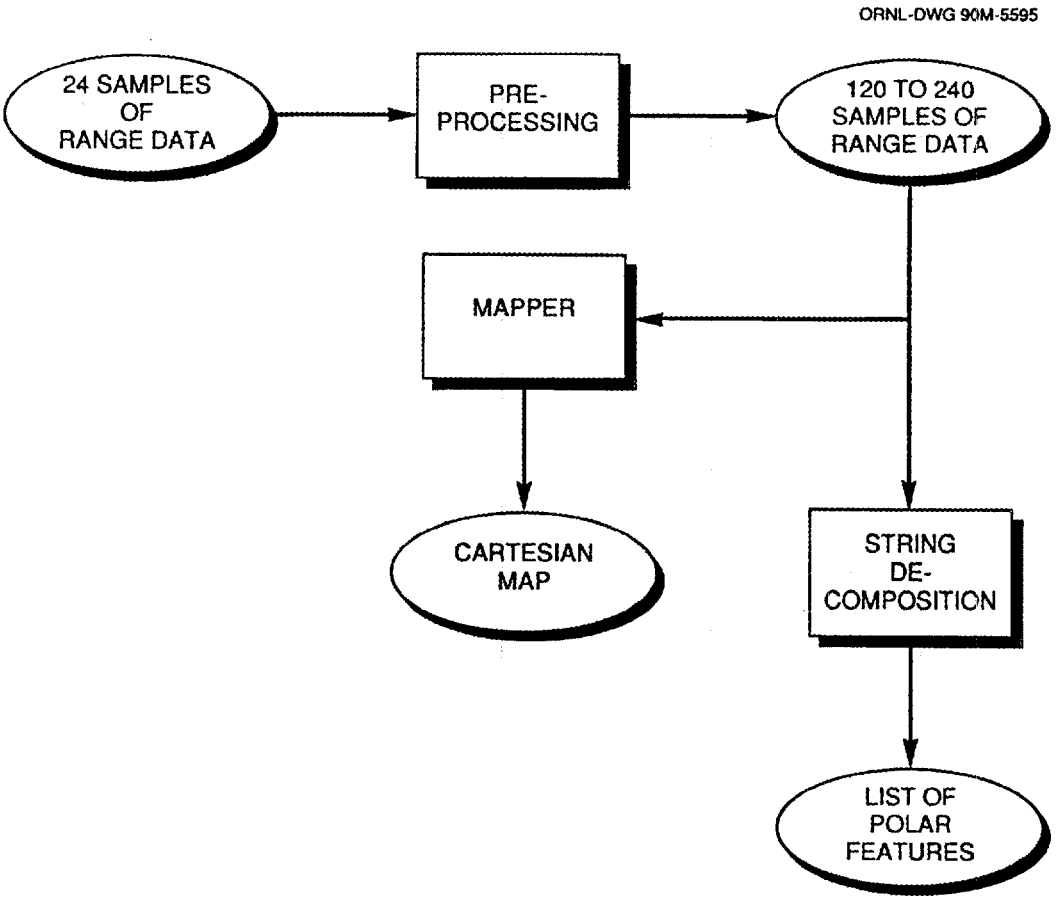

Fig. 3. Block diagram of the initial processing of the ultrasound data.

Two types of data structures, Cartesian maps and strings, were built in the second stage of the ultrasound data processing. The Cartesian map is a global, rectilinear representation of occupied and empty regions of space. It was built by transforming the (range, angle) information to Cartesian (x, y) coordinates. The cells of the map so found represent the scattering sources giving rise to the echoes, and were labelled as occupied. For each return a ray-tracing algorithm was executed to find and label as empty the pixels representing the regions lying between the robot and the

sources of the echoes. Regions which were either occluded or were located beyond the maximum range of the sensor were labelled as unknown.

The term "strings"[4] denotes a sequence of returns of similar range from neighboring scan angles, bounded at both ends by depth, or range, discontinuities. Strings are the interesting features in the ultrasound scene, marking the possible locations of object surfaces. This second type of data structure is a local, polar representation of the data, containing information on (i) the ranges and angles of the boundaries of the strings, and (ii) the (x,y) Cartesian map coordinates of each string element. For the simple configurations studied herein the depth discontinuities were unambiguously detected using a threshold odometer.

A 64 cell × 64 cell Cartesian map produced at the termination of the initial processing of a 360 deg scan is displayed in Fig. 4. Each cell of the map denotes a 15 cm × 15 cm region of space, and is labeled as either occupied, empty or unknown. We see that there are several empty zones fanning out from the robot origin. These empty zones are terminated by sequences of occupied pixels, that is, by the strings. The strings shown in the figure denote the positions of the exposed surfaces of three boxes, which had been placed on the floor, and a portion of the laboratory walls.

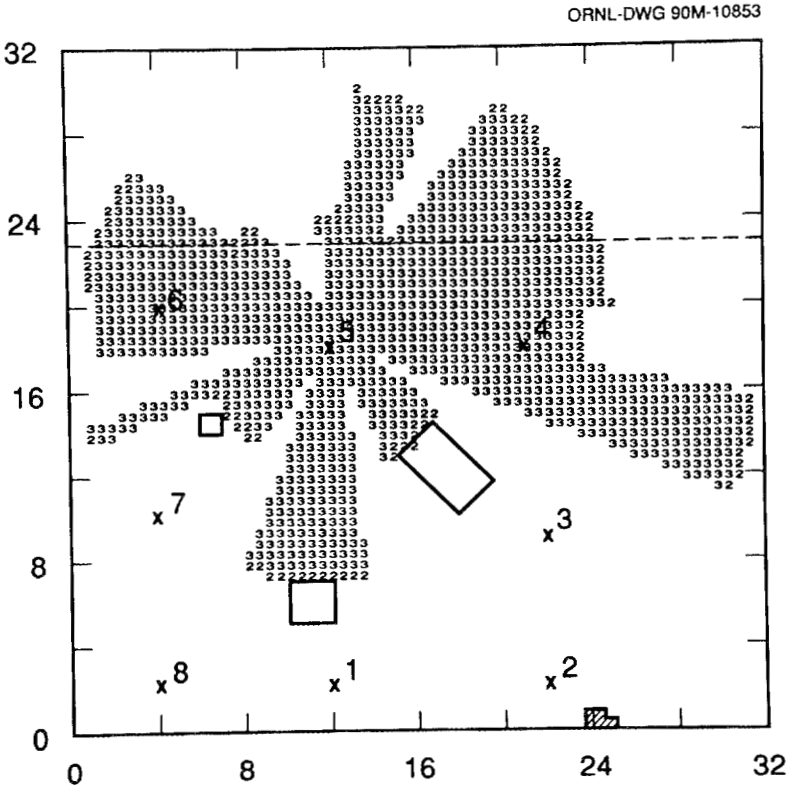

Fig. 4. Cartesian map of a setup containing three objects (boxes). Each cell denotes a 15 cm × 15 cm region of space. Cells labeled by '3's are empty; those marks by '2's are occupied, and cells left blank represent unknown areas. The robot origin is at the center of the "fan" pattern.

In constructing this figure, the labelling of empty pixels at scan angles for which no echoes was received (long returns) was suppressed. This was done in order to clearly outline the strings and their defining depth discontinuities.

The Cartesian map produced from the one box configuration shown in Fig. 1 is presented in Fig. 5. Long returns were not suppressed, and the fan pattern is less pronounced than in Fig. 4. In [4] a numerical representation was used together with a simple logic for label multiplication to mark and update the cells of the Cartesian maps. We did not perform these operations in this study, and have now changed the appearance of the displayed map in order to more easily exhibit the modifications due to fusion with the visual data. The cells labelled as empty and occluded represent the same physical dimensions as in Fig. 4. The cells labelled as occupied have a slightly different interpretation. The sizes of the occupied cells have no direct meaning. Instead, their positions, in 1.5 deg intervals, denote the (x,y) coordinates of the individual string elements.

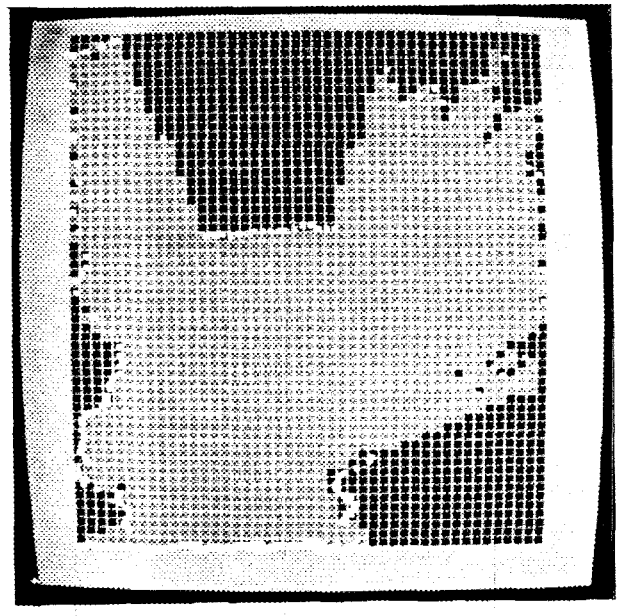

Fig. 5. Cartesian map of a the setup in Fig. 1. Empty (grey) and unknown (black) cells have the same meaning as in Fig. 4. The small (white) squares denote the string elements, that is, occupied locations in 1.5 deg steps.

3.2. Initial Processing of the Visual Image

The minimal processing of the visual image as outlined in the block diagram, Fig. 6, was straightforward. In order to obtain the binary image shown in Fig. 2, the grey-scale CCD image was convoluted with an anisotropic Gaussian and differentiated in the horizontal direction. The vertical edge segments were then identified using non-maximal suppression and local extremum finding.

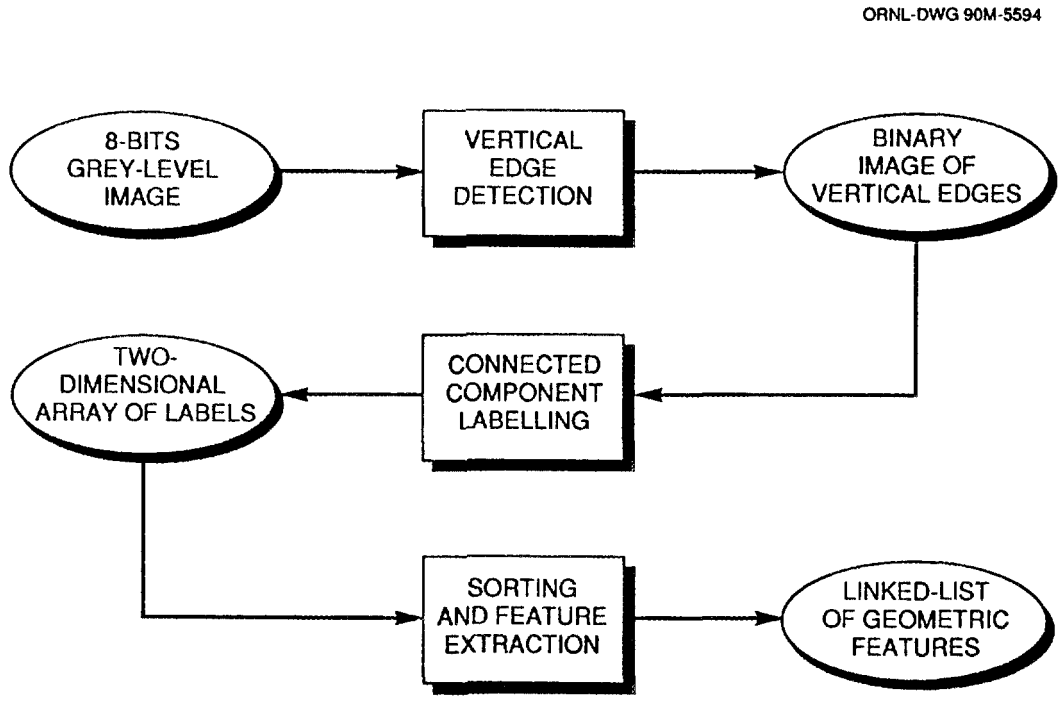

Fig. 6. Block diagram of the initial processing of the visual image.

The anisotropic Gaussian convolutions were accomplished through repeated operation with spatially-limited, additive filters. Two filtering operations were done, each using a "triangular" filter with a small number of repetitions, one along the direction of the edges, and one across the edges. This approach was based upon the observation that repeated convolution with any spatially-limited filter tends to an equivalent Gaussian convolution. Simple averaging was also tried, and found to produce results similar to that obtained using the filters. It may be noted that the maxima in the output of this operator corresponds to the zero-crossings in the output of the Laplacian, when restricted to one dimension.

The next stage of processing consisted of using a connected component algorithm to identify sequences of contiguous edge segments in the binary image, and to assign a unique label to the pixels in each sequence. The definition of contiguous was a strict one, namely, two edge segments were contiguous if they were separated from one another by no more than one pixel in the horizontal and vertical directions. That is, the edge segments were contiguous if they were 8-connected. The output of the connected component algorithm was a two-dimensional array of the assigned edge segment labels.

This array of edge segment labels was then used to build a linked-list containing the geometric features of each of the edge segments. The first part of each entry in the linked-list specified the label, the horizontal (x) coordinate, and the top and bottom vertical coordinates, ytop and ybot, of each segment. The second part of each entry in the linked-list contained pointers to the entries for the next and adjacent edge

segments of the same label. These were found by sorting the entries for each set of edge segments sharing a given label.

The final operation prior to fusion was that of weak thresholding with respect to the total length of the labelled sequences. The interesting edge segments cannot be uniquely distinguished from uninteresting ones by simple geometric properties or by their grey-scale intensity gradients. For these reasons only weak thresholding operations were carried out in the image processing. To produce Fig. 2 a weak thresholding, or non-maximal suppression, was done with respect to the magnitude of the intensity gradient. The weak thresholding operation done on total length of the labelled sequences served to remove short, isolated edge segments from the image. The results are presented in Fig. 7.

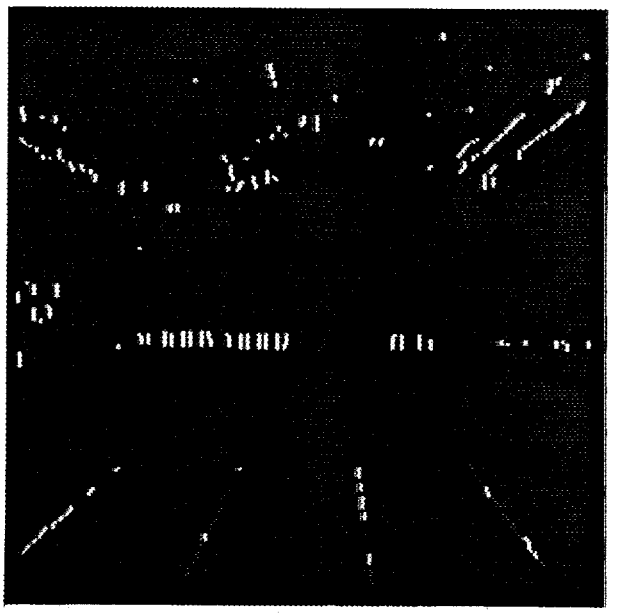

Fig. 7. Vertical edge segments after thresholding. Edge segments removed by the thresholding operation appear as short bright lines.

4. SENSOR FUSION

4.1. Fusion and Feedback

Upon completion of the initial processing, the partially interpreted sensor data were fused. This was done by exchanging information between the two sensor domains. Underlying this exchange was the observation that while the ranges extracted from the ultrasound data were intrinsically accurate to 0.25 cm, the resolution in the horizontal plane was poor. The visual edge data, on the other hand, yielded accurate horizontal distance measures when the appropriate edges were identified, that is, when the missing depth (range) perception was provided.

The mutually beneficial exchange of information between ultrasound and visual sensor domains was handled by the pair of sequentially executed "feedback" loops shown in the flow chart, Fig. 8. In the first loop, ultrasound range information was used to identify candidate vertical edge pairs. In the second loop, the resulting visual edge information was used to reduce the influences of ultrasound distortion errors upon the strings and Cartesian maps.

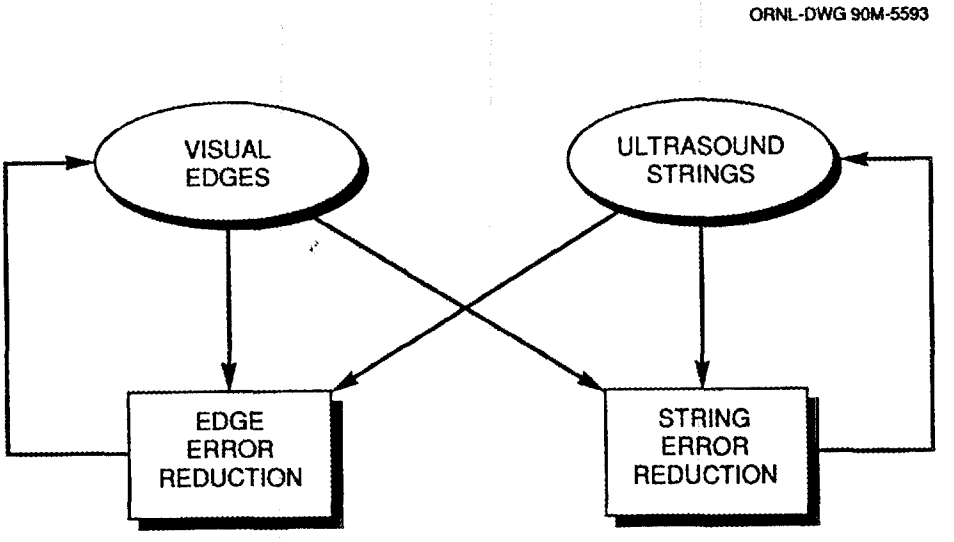

Fig. 8. Conceptual block diagram of the fusion process.

4.2. Ultrasound Information in the Visual Domain

The fusion process was initiated by using the shared horizontal angular coordinate to establish a preliminary correspondence between ultrasound strings and visual edge segments. The list of strings was first pruned of all entries whose angular extent lies outside the field-of-view of the CCD camera. The remaining strings were mapped into the visual domain. Results of this mapping are shown in Fig. 9. In this figure we observe the vertical edge segments after the weak thresholding with respect to segment length and intensity gradient magnitude. Also shown in the figure are a series of small squares. These squares are spaced in 1.5 degree intervals, and are in

a one-to-one correspondence with the ultrasound string elements displayed in Fig. 5.

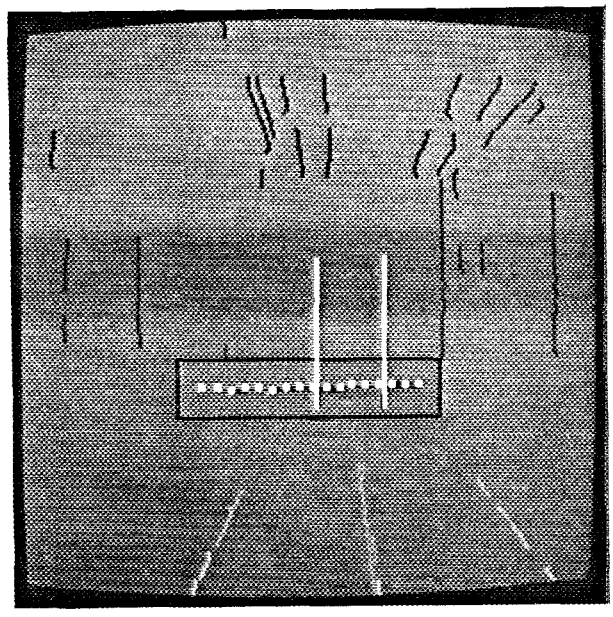

Fig. 9. Dual visual representation: Vertical edge segments and ultrasound string elements. The bounding region is indicated by the rectangle.

Returning to Fig. 1 we note that the visual picture contains a horizon which divides the scene into upper and lower domains. The lower domain is dominated by a single surface, the floor (ground), attached to which is the box we wish to identify. As can be seen in Fig. 2, the mapping of the floor coordinates to the corresponding pixel positions in the lower half of the visual image was not linear. Instead, there were distortions due to perspective geometry and the wide-angle lens of the CCD camera. To map the ultrasound strings into the visual domain a look-up table was constructed from detailed floor calibration data. This information was used together with the range data to locate the bottoms of the objects in the image. In Fig. 9 the y-coordinates of the squares denoted the location of the ground (floor) at the appropriate distance from the CCD camera.

In Fig. 9 the squares have been surrounded by a rectangular region. This bounding region represented a crude estimate of the tolerances involved in the dual string-edge representation. The bounding region was defined by the inequalities:

$$\mathrm{xmin} - \delta\mathrm{x} < \mathrm{x} < \mathrm{xmax} + \delta\mathrm{x} \qquad (1)$$

and

$$y(r) - \delta y < ybot < y(r) + \delta y \ . \tag{2}$$

In the first of these two expressions xmin and xmax denoted the string endpoints; the inequalities and large values for δx were needed in order to take into account the as yet untreated ultrasound systematic errors. In the second relation y(r) represented the y-coordinate of the floor at range, r, in the visual image, and δy was a tolerance which included uncertainties in the floor calibration and errors in detecting the bottom of the object in the environment using the vertical edge detection algorithm. These expressions provided bounds on the coordinates, x and ybot, of the sets of vertical edges of interest.

In Fig. 9 the two sequences of vertical edges falling within these bounds have been highlighted. As seen in the figure, the highlighted features correspond to the vertical edges of the box sitting on the floor of the CESAR lab. In greater detail, we note that the vertical lines accurately represent the horizontal dimensions of the box. One of the goals of the Gaussian smoothing was to reduce the sizes and numbers of the breaks in the edge segments. In accomplishing this goal the smoothing increased the lengths of the edge segments beyond the boundaries in the vertical direction. This lengthening was taken into account by δy.

4.3. Statistical Analysis

Our objective in performing a statistical analysis was to extract the salient geometric features of each extended sequence of vertical edge segments sharing a common label. In doing so we noted that many background features such as floor markings and overhead lighting fixtures gave rise to 8-connected edge sequences possessing overall non-vertical slopes. These slopes (and intercepts), when found, could be used as an additional spatial filter to eliminate the unwanted features from the image, thereby increasing the robustness of the bounding region algorithm.

The statistical analysis consisted of least-squares, linear fits plus an evaluation of the results using the calculated values for chi-square. In performing a least-squares analysis it should be noted that we do not assume that the distribution of errors is gaussian. The justification for this type of fitting resides in the observation (see, for example, [8, 31]) that least-squares fits provide the most accurate description possible of a set of data (vertical edge segments). That is, this fit corresponds to a condition of minimum error or, equivalently, to a condition of maximum weight, which for a Gaussian is also a condition of maximum likelihood.

The statistical analysis was simplified by dividing the connected sequences into two groups with respect to their distribution in (horizontal) x-coordinates. The first group contained those sequences having narrow distributions of their x-coordinates, specifically, sequences whose width was no more than two or three pixels. A full statistical analysis was unnecessary for these cases. Instead, we calculated the length-weighted first moment of the distribution of x-coordinates:

$$< x > = \frac{\sum_i x_i \lambda_i}{\sum_i \lambda_i} \tag{3}$$

where λ_i is the length of the ith segment. The original sequence of vertical edge segments was replaced by a single line located at $< x >$, extending from the top to the bottom of the sequence. The bright lines displayed in Fig. 8 were generated in this manner.

The second group contained all other groups of extended sequences of vertical edge segments. The method adopted by us for fitting a least-squares line to a set of vertical edge segments sharing a common label was to use the midpoints of each segment as the (vertical) y-coordinate, and weigh the corresponding term by the length of that segment. This simple procedure can be justified by noting that for any connected sequence of edge segments the excursions in the horizontal direction were just one pixel per segment. This ansate was also in accordance with our intuitive notion that long edge segments are far more likely to correspond to actual depth discontinuities than short edge segments.

The y-coordinate midpoints were treated as the dependent variable and the x-coordinate as the independent variable in the least-squares fits. The corresponding least-squares line of slope 'b' and intercept 'a' was determined from the data as

$$a = \frac{1}{\Delta}\begin{vmatrix} \sum_i \frac{y_i}{\sigma_i^2} & \sum_i \frac{x_i}{\sigma_i^2} \\ \sum_i \frac{x_i y_i}{\sigma_i^2} & \sum_i \frac{x_i^2}{\sigma_i^2} \end{vmatrix} \qquad b = \frac{1}{\Delta}\begin{vmatrix} \sum_i \frac{1}{\sigma_i^2} & \sum_i \frac{y_i}{\sigma_i^2} \\ \sum_i \frac{x_i}{\sigma_i^2} & \sum_i \frac{x_i y_i}{\sigma_i^2} \end{vmatrix} \tag{4}$$

with determinant

$$\Delta = \begin{vmatrix} \sum_i \frac{1}{\sigma_i^2} & \sum_i \frac{x_i}{\sigma_i^2} \\ \sum_i \frac{x_i}{\sigma_i^2} & \sum_i \frac{x_i^2}{\sigma_i^2} \end{vmatrix} \tag{5}$$

and weight factor

$$\sigma_i = \frac{1}{\sqrt{\lambda_i}} \tag{6}$$

Least-squares lines generated using Eqs. 4 to 6 are shown superimposed upon the appropriate sequences of edge segments in Fig. 10. These fitted lines, as well as the bright lines shown in Fig. 9, were added to the visual image using Bresenham's [5] line-drawing algorithm. As can be seen in the figure, the least- squares lines are nearly indistinguishable from the underlying connected edge segments.

In Fig. 10 the least-squares lines correspond to details of the overhead lighting fixtures. These lines were identified easily by their slopes which were small compared to those of the lines describing the vertical edges of the object of interest, i.e., the box standing on the floor of the lab. By the end of the least-squares analysis these and similar lines were removed from the list of lines using this "slope filter". The output from the analysis was a short list of lines and their geometric features. The geometric features consisted of the coordinates of one endpoint (the top), and the slope and intercept.

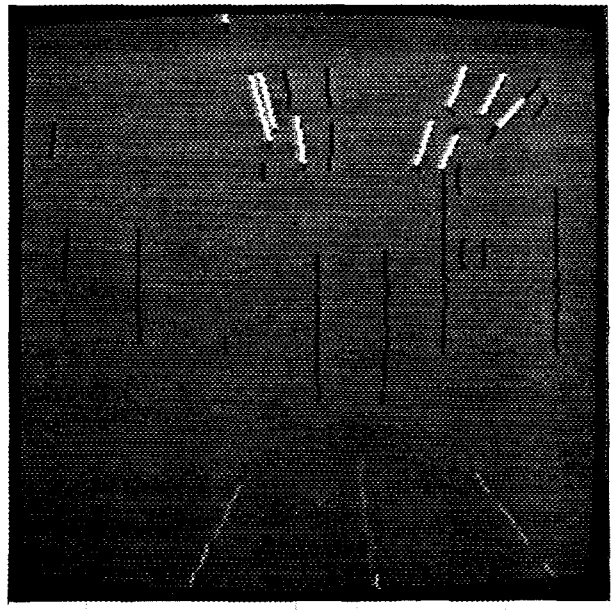

Fig. 10. Least-squares fitting. Least squares lines are shown as bright lines superimposed upon the underlying sequences of edge segments.

The final quantity of interest was the value for chi-square. We had found that, in some instances, sets of edge segments belonging to different features in the visual image were 8- connected. For example, for certain placements of the box on the floor of the lab, the floor markings were merged into the vertical edges of the box. The solution to this type of systematic error was to use the value of chi-square as a criterion for identifying their occurrence, since the quality of the fits deteriorated whenever sequences of edge segments of differing slope were merged into one another. An iterative end- point fitting algorithm [9] was applied to locate the region(s) of inflection, and the two sets of edge segments were unlinked by fitting more than one line to the data.

The expression used by us for calculating the value for chi- square was

$$\chi^2 = \sum_i \left[\frac{1}{\sigma_i^2} \left(y_i - a - bx_i \right)^2 \right] \tag{7}$$

Similarly, the reduced chi-square was defined as

$$\chi^2_{\text{red}} = \frac{\chi^2}{(L - 2)} \tag{8}$$

In Eq. 8, L denotes the total length of the edge segments in the sequence. For simplicity, the deviations were only calculated for the midpoint y-coordinates of the edge segments. This served as an adequate means for distinguishing between good and bad (erroneous) fits. To conclude this section it should be noted that Andress and Kak [2] defined similar linking, unlinking and merging operations in their production system work.

4.4. Visual Information in the Ultrasound Domain

Having discussed how the ultrasound information was incorporated into the visual image, we now turn to the second feedback process, where visual sensor data was used to modify and correct the (ultrasound) Cartesian map. The starting point for the corrections to the map was the dual representation of the ultrasound and visual data depicted in Fig. 9. This representation enabled direct comparisons of the string elements to the highlighted vertical lines. Specifically, a minimum distance algorithm was used to find the ultrasound string element nearest each of the identified visual vertical lines. These vision-defined string elements were interpreted as the corrected string endpoints.

We recall that the string data structures contain the identifying (x,y) coordinates of the corresponding occupied cells of the Cartesian map. This information was used to modify the Cartesian map to reflect the corrections to the string endpoints. The results are exhibited in Fig. 11. Upon comparison of the two maps, Figs. 5 and 11, we find that the exposed surface of the box has been modified to reflect the visual information. The remainder of the map was left unchanged. In modifying the Cartesian map, cells at the appropriate scan angles initially labelled as either occupied or occluded were relabelled as empty. The ranges assigned at these angles were based upon data from scan angles lying just beyond the original string endpoints.

It is worthwhile to note that the simple combination of ultrasound and visual sensor data yielded estimates of linear dimensions of objects which were both accurate and precise. For example, the ultrasound range estimate for the box was 4.42 m. In Fig. 9 the ultrasound string data provided an absolute scale for the visual edge angular separation. Upon combining this information with the range data we obtained a box width of 59.8 cm. The actual width of the box was 61.0 cm; that is, our result was accurate to 2%. The accuracy declined as the yaw of the box increased. In those situations the strings tended to bend away from the surface of the box and the range estimates denoted the distance to the nearest part of the box only. The outer (diagonal) edges were imaged, while the interior edge of the box became progressively more difficult to observe.

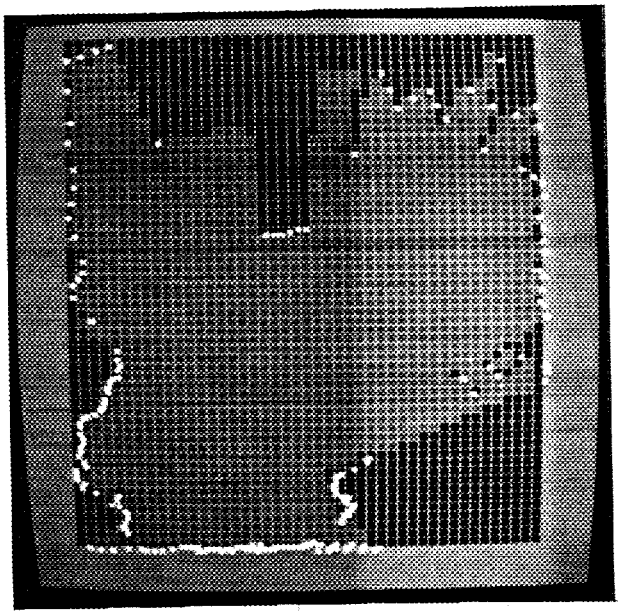

Fig. 11. Cartesian map of the setup shown in Fig. 1, after sensor fusion. Symbols have the same meaning as in Fig. 5.

Turning to the precision of the width estimates, we note that the representation of visual and ultrasound data in Fig. 9 yielded far more precise values for widths than did the Cartesian map, Fig. 11. The intrinsic precision of Cartesian map was given by the 15 cm cell size. The dual representation was a factor of 5 more precise. To obtain the overall precision of the 59.8 cm width estimate, we combined the ultrasound range and visual pixel errors in quadrature to yield an error of 3.2 cm.

V. SUMMARY AND CONCLUDING REMARKS

To summarize, we have presented a methodology for fusing ultrasound and visual sensor data as acquired by a mobile robot. The objective of the fusion process was to reduce the systematic errors which arise in the processing of the data in the individual ultrasound and visual sensor domains. The need for strategies of this type arises whenever autonomous robots, or other automated data processing systems, begin to interpret data collected in environments which are unstructured to any appreciable extent.

We noted both in the present study and in our previous one on ultrasound data fusion that there were two distinct error treatment stages. In the identification stage comparisons were made between different sets of initially interpreted sensor data. In the resolution stage, models of the sensors, the environment and their mutual interactions were used to modify the initial interpretations and remove the errors flagged in the first stage. This two stage method for error treatment is analogous to that of a feedback process.

In the ultrasound domain examined herein we focussed our attention on the distortion errors produced by the limited angular resolution of the beam. In the visual domain we extracted vertical edge information, and the systematic errors involved the well-known ambiguities in discriminating depth discontinuities from intensity gradients generated by other details in the image. To flag these sets of errors we compared the ultrasound strings, that is, polar representations of the ultrasound scan data, and visual vertical edges to one another. The comparisons were straightforward to do, since the data structures in the two sensor domains shared a common angular coordinate. By the end of the second stage, the data structures in each sensor domain carried some information from the other domain. We found that we could identify the vertical edges of interest, tag them with range information, and remove the distortions in size from the Cartesian navigation maps.

From the viewpoint of the visual sensor domain, a series of operations was applied to the visual image to extract the vertical edges corresponding to depth discontinuities. The edges of interest were neither intense nor prominent, and could not be extracted using low-level thresholding and morphological operations, alone. Instead, we made use of the ultrasound information, the special horizontal orientation of the camera, and statistical techniques to perform additional spatial filtering of the visual data. Likewise, from the viewpoint of the ultrasound sensor domain, a series of non-local algorithms were applied to the processed data to remove distortion errors. The operation of snipping back the length of the strings to match that of the visual edge separation was formally equivalent to the pattern analysis done in [4] involving two strings, and represents an additional spatial filter for the ultrasound data.

Acknowledgment

The authors wish to thank Judson P. Jones for the many valuable discussions as well as for providing technical assistance. The authors also wish to acknowledge the support of R. C. Mann throughout the course of this study.

References

[1] Allen, P. K., "Object Recognition using Vision and Touch," Ph.D. Thesis, Department of Computer and Information Science, Univ. of Penn., Philadelphia, PA, 1985.

[2] Andress, K. M. and A. C. Kak, "A Production System Environment for Integrating Knowledge with Vision Data," *Proceedings of the 1987 Workshop on Spatial Reasoning and MultiSensor Fusion*, Ed. A. C. Kak and S. Chen, pp. 1-12, 1987.

[3] Ayache, N. and O. D. Faugeras, "Building, Registering and Fusing Noisy Visual Maps," *Int. J. Robotics. Res.*, **6**, pp. 45-65, (1988).

[4] Beckerman, M. and E. M. Oblow, "Treatment of Systematic Errors in the Processing of Wide Angle Sonar Sensor Data for Robotic Navigation," *IEEE Transactions Robotics and Automation*, Vol. RA-6, pp. 137–145 (1990).

[5] Bresenham, J. E., "Algorithm for Computer Control of Digital Plotter," *IBM Sys. J.*, **4**, pp. 25-30, (1965).

[6] Canny, J., "A Computational Approach to Edge Detection," *IEEE Transactions Pattern Anal. Machine Intel.*, Vol. PAMI-8, pp. 766-778 (1986).

[7] Chatila, R. and J.-P. Laumond, "Position Referencing and Consistent World Modelling for Mobile Robots," *Proceedings IEEE International Conference on Robotics and Automation*, pp. 138-145 (1985).

[8] Cohen, E. R., "The Basis for the Criterion of Least Squares," *Rev. Mod. Phys.*, 25, pp. 709-713 (1953).

[9] Duda, R. O. and P. E. Hart, *Pattern Classification and Scene Analysis*, New York: John Wiley & Sons, 1973, pp. 338-339.

[10] Duda, R. O., D. Nitzan and P. Barrett, "Use of Range and Reflectance Data to Find Planar Surface Regions," *IEEE Transactions Pattern Anal. Machine Intel.*, Vol. PAMI-1, pp. 259-271 (1979).

[11] Elfes, A., "Sonar-Based Real-World Mapping and Navigation," *IEEE Journal of Robotics and Automation*, Vol. RA-3, pp. 249-265 (1987).

[12] Flynn, A. M., "Redundant Sensors for Mobile Robot Control," M.S. Thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA. (1985).

[13] Gibson, J. J., *The Perception of the Visual World*, Boston: Houghton Mifflin (1950).

[14] Haber, R. N. and M. Hershenson, *The Psychology of Visual Perception*, New York: Holt, Rinehart and Winston, 2nd. ed., (1980), pp 185-191.

[15] Haralick, R. M., "Digital Step Edges from Zero Crossing of Second Directional Derivatives," *IEEE Transactions Pattern Anal. Machine Intel.*, Vol. PAMI-6, pp. 58-68, (1984).

[16] Hu, G. and G. Stockman, "3-D Scene Analysis via Fusion of Light Striped Image and Intensity Image," *Proceedings of the AAAI Workshop on Spatial Reasoning and Multi-Sensor Fusion*, A. Kak and S. Chen, ed., Morgan Karfman, Los Altos, Ca, pp. 138-147 (1987).

[17] Hueckel, M. H., "An Operator which Locates Edges in Digitized Pictures," *J. Ass. Comput. Mach.*, **18**, pp. 113-125 (1971).

[18] Killough, S. M. and W. R. Hamel, "Sensor Capabilities for the HERMIES Experimental Robot," *Proceedings of the ANS Third Topical Meeting on Robotics and Remote Systems*, Sec 4.1, pp. 1-7 (1989).

[19] Magee, M. J., B. A. Boyter, C. H. Chien and J. K. Aggarwal, "Experiments in Intensity Guided Range Sensing Recognition of Three-Dimensional Objects," *IEEE Transactions Pattern Anal. Machine Intel.*, Vol. PAMI-7, pp. 629-637 (1985).

[20] Marr, D. and E. Hildreth, "Theory of Edge Detection," *Proc. Roy. Soc. Lond.*, Vol. B 207, pp. 187-217 (1980).

[21] McKendall, R. and M. Mintz, "Robust Fusion of Location Information," *Proceedings IEEE International Conference on Robotics and Automation*, pp. 1239-1244 (1988).

[22] Nalwa, V. S. and T. O. Binford, "On Detecting Edges," *IEEE Transactions Pattern Anal. Machine Intel.*, Vol. PAMI-8, pp. 699-714 (1986).

[23] Nandhakumar, N. and J. K. Aggarwal, "Integrated Analysis of Thermal and Visual Images for Scene Interpretation," *IEEE Transactions Pattern Anal. Machine Intel.*, Vol. PAMI-10, pp. 469-481 (1988).

[24] Nevatia, R. and K. R. Babu, "Linear Feature Extraction and Description," *Comput. Graphics Image Processing*, **13**, pp. 257-269 (1980).

[25] Nitzan, D., A. E. Brain and R. O. Duda, "Measurement and Use of Registered Reflectance and Range Data in Scene Analysis," *Proceedings of the IEEE*, **65**, pp. 206-220 (1977).

[26] Prewett, J., "Object Enhancement and Extraction," in Picture Processing and Psychopictorics, B. Lipkin and A. Rosenfeld, Eds., New York: Academic Press, pp. 75-149 (1970).

[27] Roberts, L. G., "Machine Perception of Three-Dimensional Solids," *Optical and Electro-Optical Information Processing*, J. T. Trippett, et al., Eds., Cambridge, MA: MIT Press, pp. 159-197 (1965).

[28] Rosenfeld, A. and M. Thurston, "Edge and Curve Detection for Visual Scene Analysis," *IEEE Trans. Computers*, Vol. C-20, pp. 562-569 (1971).

[29] Stanfield, S., "Visually-Guided Haptic Object Recognition," Ph.D. Thesis, Department of Computer and Information Science, Univ. of Penn., Philadelphia, PA, (1987).

[30] Wang, Y. F. and J. K. Aggarwal, "On Modelling 3-D Objects Using Multiple Sensory Data," *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1098-1103 (1987).

[31] Wittiker, E. and G. Robinson, *The Calculus of Observations*, 4th ed., London: Blackie and Son, 1944, pp. 209, 224.

[32] Yuille, A. L. and T. A. Poggio, "Scaling Theorems for Zero Crossings," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-8, pp. 15-25 (1986).

ORNL/TM-11349
CESAR-89/38

Internal Distribution

1. B. R. Appleton
2. D. L. Barnett
3-7. M. Beckerman
8. B. L. Burks
9. P. L. Butler
10. G. de Saussure
11. J. R. Einstein
12-16. L. A. Farkas
17. C. W. Glover
18. W. R. Hamel
19. J. N. Herndon
20-24. S. E. Johnston
25. J. P. Jones
26. H. E. Knee
27. C. T. Kring
28. G. Liepins
29-33. F. C. Maienschein
34-38. R. C. Mann
39. S. A. Meacham
40. E. M. Oblow
41-45. F. G. Pin
46. P. F. Spelt
47. F. J. Sweeney
48. J. J. Dorning (consultant)
49. R. M. Haralick (consultant)
50. J. E. Leiss (consultant)
51. N. Moray (consultant)
52. M. F. Wheeler (consultant)
53. EPMD Reports Office
54. Central Research Library
55. ORNL Technical Library
Document Reference Section
56-57. Laboratory Records
58. ORNL Patent Office
59. Laboratory Records-RC

External Distribution

60. Office of the Assistant Manager, Energy Research and Development, DOE-ORO, Oak Ridge, TN 37831

61. Dr. Peter Allen, Department of Computer Science, 450 Computer Science, Columbia University, New York, NY 10027.

62. Dr. Harry Alter, Office of Technology Support Programs, Office of Nuclear Energy, U. S. Department of Energy, Washington, DC 20545

63. Dr. Ruzena Bajcsy, Department of Computer and Information Science, University of Pennsylvania, 3401 Walnut Street, Philadelphia, PA 19104

64. Dr. Wayne Book, Department of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332

65. Dr. Rodney Brooks, MIT Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139

66. Dr. J. S. Coleman, Division of Engineering, Mathematical and Geosciences, Office of Basic Energy Sciences, ER-15, U. S. Department of Energy – Germantown, Washington, DC 20545

67. Dr. Ewald Heer, Heer Associates, 5329 Crown Avenue, La Canada, CA 91011

68. Dr. Avi Kak, Robot Vision Lab, School of Electrical Engineering, Purdue University, West Lafayette, IN 47907

69. Dr. Tomas Lozano-Perez, MIT Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139

70. Dr. Oscar P. Manley, Division of Engineering, Mathematical and Geosciences, Office of Basic Energy Sciences, ER-15, U. S. Department of Energy - Germantown, Washington, DC 20545

71. Dr. Max Mintz, Department of Computer and Information Science, University of Pennsylvania, 3401 Walnut Street, Philadelphia, PA 19104

72. Dr. Tom Mitchell, The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213

73. Dr. Hans Moravec, The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213

74. Dr. Michael J. Rabins, Department of Mechanical Engineering, Texas A & M University, 100 Engineering Physics Building, College Station, TX 77843

75. Dr. Karl N. Reid, College of Engineering, Architecture and Technology, Oklahoma State University, Stillwater, OK 74074

76. Dr. Melvin W. Siegel, The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213

77–86. Office of Scientific and Technical Information, P. O. Box 62, Oak Ridge, TN 37831