ıs using the heat equa-
'rn Recognition. IEEE

ıon and regularization

visual surface recon-
(1983).
ıtion of ill-posed prob-
–89 (1987).
of AAAI-87: Sixth Na-
Publishers, 1987, pp.

ıtice Hall, Englewood

and Macromolecules:
son-Wesley, Reading,

. 1983.
tereo," Ph.D. thesis,

search using dynamic
l., **PAMI-7**, 139–154

ssive sensing technol-
, N nber 1989, pp.

ler Wal, "A pipelined
sion, NATO ASI Se-
52.
Comp. Graph. Image

lifference of low-pass
II-6, 212–222 (1984).
etectability in passive
r Vision and Pattern

ile robots," technical

anning and execution
ational Conf. on Ro-
y 1990, pp. 20–25.

Inertial Navigation Sensor Integrated Motion Analysis for Autonomous Vehicle Navigation

Barry Roberts and Bir Bhanu*
Honeywell Systems and Research Center
Minneapolis, MN 55418

Received February 10, 1992; accepted April 5, 1992

Many types of existing vehicles contain an inertial navigation system (INS) that can be utilized to greatly improve the performance of motion analysis techniques and make them useful for practical military and civilian applications. This article presents the results obtained with a maximally passive system of obstacle detection for ground-based vehicles and rotorcraft. Automatic detection of these obstacles and the necessary guidance and control actions triggered by such detection will facilitate autonomous vehicle navigation. Our approach to obstacle detection employs motion analysis of imagery collected by a passive sensor during vehicle travel to generate range measurements to world points within the field of view of the sensor. The approach makes use of INS data and scene analysis results to improve interest point selection, the matching of the interest points, and the subsequent motion-based range computations, tracking, and obstacle detection. In this article, we concentrate on the results obtained using lab and outdoor imagery. The range measurements that are made by INS integrated motion analysis are compared to a limited amount of ground truth that is available. © 1992 John Wiley & Sons, Inc.

現在、多くのタイプの車両が慣性航法システム（INS）を採用しており、運動解析技術の大幅な改善に役立っている。そして、このシステムは実際の軍事用と民間用の車両で活用されている。ここでは、地上用車両と回転翌機用に障害物を検出する最大受動システムから得られた結果について解説する。障害物の自動検出とこれによって発生する必要な誘導と制御は、車両の自動運行を支援する。今回の障害物検出システムでは、移動中車両の受動センサーで収集した画像による運動解析を採用し、センサーの視野におけるワールド・ポイントのレンジ測定を行う。このINSデータの活用によって、対象ポイント選択、対象ポイントのマッチング、連続運動検出、トラッキング、障害物検出を改善している。さらに、実験室と野外で採取した画像データから得られた結果を一つにまとめて説明している。また、INSの集中運動解析によるレンジ測定は、限られた数の地上の基準点と比較される。

*Present address: College of Engineering, University of California, Riverside, CA 92521.

Journal of Robotic Systems 9(6), 817–842 (1992).
CCC 0741-2223/92/060817-026$04.00

1. INTRODUCTION

A variety of active sensor-based techniques for obstacle detection have been explored to date.[1-3] These approaches mainly focus on the processing of laser range (ladar) imagery and millimeter wave (MMW) radar data. In our approach, we prefer a passive sensor that will enable the vehicle to be covert and therefore minimize any possible threat to the vehicle and the pilot. Of equal importance are the field of view, resolution of the data used for obstacle detection, and the access time of such data. Both MMW and ladar suffer in one of these categories.

Passive sensors, such as a TV camera, are also being used to detect obstacles for ground vehicles.[4-7] However, state-of-the-art motion analysis techniques for obstacle detection are not robust and reliable enough for many practical applications. Many of these techniques require that unrealistic constraints be placed on the input data to make them work. The largest sources of errors are unknown sensor motion and incomplete/ambiguous information in the sensed image data. However, many types of land and air vehicles contain an inertial navigation system (INS) whose output can be used for applications beyond the original intent of the system. Within such vehicles, the INS information can be used to greatly simplify some of the functionalities normally provided by computer vision, such as obstacle detection, target motion detection, target tracking, etc.

In this article, we describe the use of INS measurements to enhance the quality and robustness of motion analysis techniques for obstacle detection and thereby provide vehicles with new functionality and capability. The objective of the work presented in this article is to present the results of our obstacle detection approach when applied to sequences of indoor (laboratory) and outdoor imagery for which synchronized INS data exist.

Before entering into the technical discussions of our approach to motion analysis, we will first provide the reader with a brief description of an INS and the type of data it generates. An INS includes an inertial reference unit (IRU) and all necessary hardware for stabilizing and processing the IRU outputs to derive values for the position and velocity (of whatever platform to which the INS is attached) in a desired reference frame.[8] An IRU is defined as an assembly of instruments capable of providing full 3-D measurement of absolute rotation and nongravitational acceleration. The measurements are typically made relative to a stationary nonrotating coordinate frame, with the use of gyroscopes and accelerometers. Initially, gyroscopes were constructed of a spinning mass attached to a gimballed platform whose attitude, relative to the vehicle, was measurable. Modern gyroscopes consist of inertial instruments (e.g., ring lasers) mounted to the vehicle's axes (one per axis). Such systems are referred to as *strapdown* systems due to their lack of movement relative to the vehicle. The modern gyroscope has its output stabilized computationally (with the aid of computers) instead of mechanically. Regardless of the method of implementation, the gyroscope provides an absolute measure of the rotation difference between the vehicle's coordinate frame and a fixed, geographic, reference frame. Such knowledge of the vehicle's *attitude* is important in processing the measurements made by accelerometers (one per vehicle axis).

Accelerometers make measurements of vehicle acceleration along the three vehicle axes. Hence, to convert the acceleration measurements such that they are relative to the fixed, reference coordinate frame, the knowledge of the vehicle's attitude relative to the reference frame is crucial, for obvious reasons. With the knowledge of vehicle acceleration relative to the reference frame, time integrals of the acceleration measurements are performed to generate values of vehicle velocity and position. Of particular interest to this article are the measurements of vehicle velocity, $\bar{v}$, and attitude, (ϕ, θ, ψ). The exact use of $\bar{v}$ and (ϕ, θ, ψ) will be described in detail later in the article.

In Section 2, we briefly review our approach to motion analysis by describing the fundamental details of the approach. Section 3 describes the results we have obtained with our INS integrated motion analysis algorithm. Finally, Section 4 provides the conclusions.

2. INERTIAL SENSOR INTEGRATED MOTION ANALYSIS

The purpose of this section is to describe the inertial sensor integrated motion analysis approach we have undertaken. The block diagram of this system is illustrated in Figure 1. The system uses inertial sensor integrated motion analysis, scene analysis, and selective applications of active sensors to provide an obstacle detection capability.[4]

As shown in Figure 2, the data input to the obstacle detection algorithm consists of a sequence of digitized video or FLIR frames that are accompanied by inertial data consisting of rotational and translational velocities. This infor-

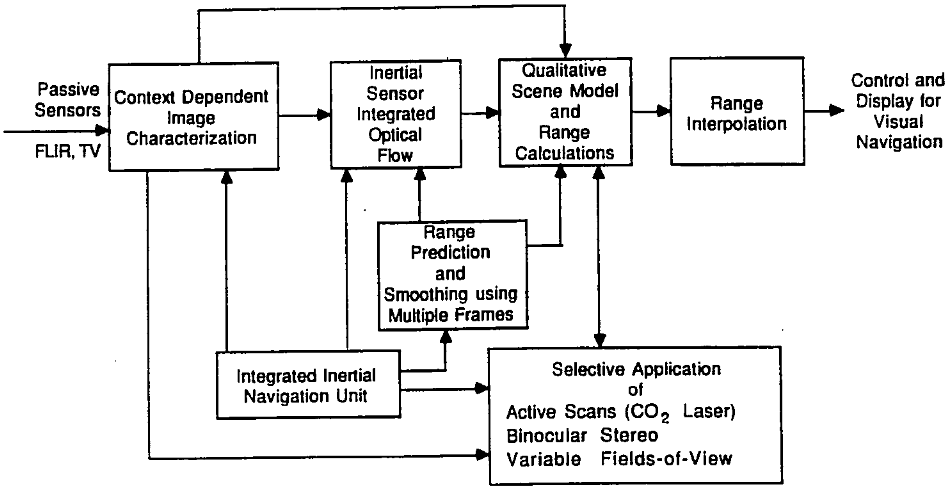

Passive Sensors
FLIR, TV
Context Dependent Image Characterization
Inertial Sensor Integrated Optical Flow
Qualitative Scene Model and Range Calculations
Range Interpolation
Control and Display for Visual Navigation
Range Prediction and Smoothing using Multiple Frames
Integrated Inertial Navigation Unit
Selective Application of Active Scans (CO2 Laser) Binocular Stereo Variable Fields-of-View

Figure 1. Inertial sensor integrated motion and scene analysis using both passive and selective applications of active sensors provide robust image analysis useful for obstacle detection/avoidance by a robotic land vehicle or helicopter.

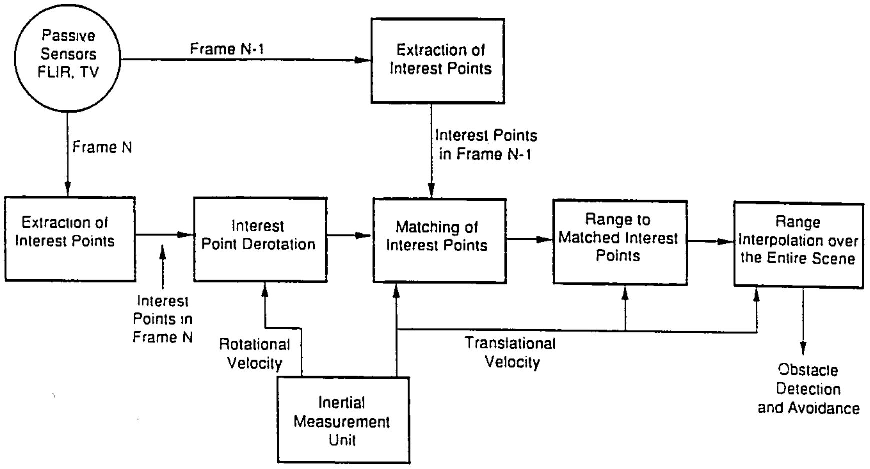

Passive Sensors FLIR, TV
Frame N-1
Extraction of Interest Points
Frame N
Interest Points in Frame N-1
Extraction of Interest Points
Interest Point Derotation
Matching of Interest Points
Range to Matched Interest Points
Range Interpolation over the Entire Scene
Interest Points in Frame N
Rotational Velocity
Translational Velocity
Inertial Measurement Unit
Obstacle Detection and Avoidance

Figure 2. Inertial sensor integrated motion analysis technique.

mation, coupled with the temporal sampling interval between frames, is used to compute the distance vector, $\vec{d}$, between each pair of frames and the roll, pitch, and yaw angles, (ϕ, θ, ψ), of each frame. Both $\vec{d}$ and (ϕ, θ, ψ) are crucial to the success of the algorithm, as will be described later.

The blocks shown in Figure 2 define the major steps involved within the *obstacle detection using inertial navigation data* (*ODIN*) motion analysis algorithm suite. In the subsections that follow, we briefly address the function of these blocks.

Before starting a detailed discussion of the major steps in the algorithm, let us first describe the coordinate systems that are used. The digitized imagery contains pixels addressed by row and column with the origin of the 2-D coordinate system located in the upper left corner of the image. The horizontal axis, c, points to the right and the vertical axis, r, is in the downward direction. This image plane is perpendicular to the x-axis of a 3-D coordinate system and is located at a distance of the focal length, F, from the origin with the z-axis in the downward direction. Therefore, the pixels in the image plane can be described in the 2-D coordinate frame as (c, r) and in the 3-D coordinate frame by the vector (F, y, z). The geometry described above is graphically illustrated in Figure 3. With knowledge of the sensor field of view and F, the transformation between (c, r) and (F, y, z) is easily computed.

2.1. Distinguishing Features

The features within the imagery (TV or FLIR) that are most prominent and distinguishing mark the world points to which range measurements will be made. These prominent world points, known as *interest points*, are (by definition) those points that have the highest promise of repeated extraction through-

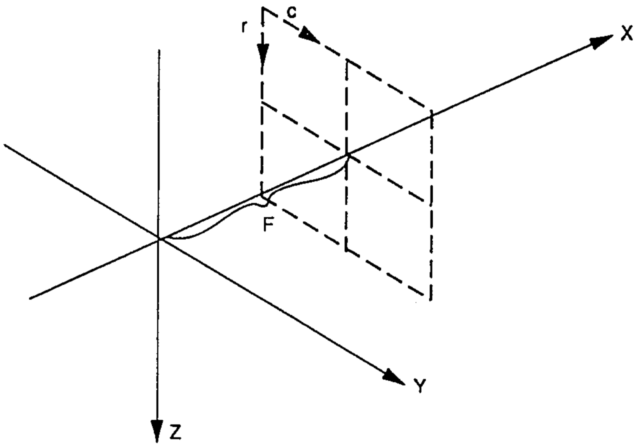


Figure 3. The coordinate system geometry of the sensor's image plane is perpendicular to the x-axis, located at the distance of the focal length, F, from the origin of the coordinate system.

out multiple frames. The interest points within the field of view of the monocular sensor are of fundamental and critical importance to motion analysis calculations. In the following subsections, the extraction of interest points is described.

2.1.1. Image Segmentation

Unfortunately, not all regions within a scene can contain reliable interest points. Hence, we employ scene analysis techniques to ascertain a measure of *goodness* for each region prior to interest point selection.[9] The resulting interest point extraction routine takes as input a segmentation of the original image and returns n_j, $0 \leq j \leq N$, interest points in each of the N segments. The value of n_j for segment j is proportional to the segment size and other segment features. More than n_j interest points can exist per segment; only the points with the highest *interestingness* values are reported. The result of incorporating scene segmentation results into interest point extraction is that, for a given scene, the interest points are more uniformly distributed.

The outcome of the two-phase segmentation procedure, *initial segmentation* (based upon image texture) and *region merging*, is the allocation of a number of interest points to each region. This number is indicated by a percentage of the total number of interesting points for the whole image that are desired within each individual region. In assigning the region percentages we use edge information and the gray value mean and variance within each region. Regions that have very high mean are indicative of sky regions in which interest points are few unless wires and poles protrude into the sky. For regions of high mean, the presence of edges is of paramount importance in computing the desired number of interest points. If the number of edge pixels in sky regions is sufficiently high, the region's gray scale variance and area (in terms of pixels) are used to compute the region's percentage, i.e.,

$$\text{region percentage} = 100\left(w_a \frac{\text{area}_i}{\text{area}_{\text{total}}} + w_v \frac{\text{variance}_i}{\text{variance}_{\text{total}}}\right)$$

Region percentage is computed similarly for nonsky regions. It is important to note that the region segmentation is intentionally kept coarse; it would be counterproductive to reduce region size to the point where regions are so small that each region is homogeneous in gray level. Such homogeneous regions would be featureless.

Scene segmentation is also used in range point interpolation/surface fitting. Range to interest points that all lie within one region will be interpolated together to generate one surface. All of the resulting range surfaces belonging to the various regions will be joined together at their respective boundaries to form one continuous surface.

The scene segmentation procedures requires several processing steps that can be divided into two groups, *initial segmentation* and *region merging*, as described below.

Initial Segmentation

1. Compute the local mean "image" and the local standard deviation image of the input image. The local region used to compute mean and standard deviation is a 7 × 7 window.
2. Compute the *texture gradient* image of the input image using the mean and standard deviation images. The texture gradient is the maximum of four measurements made with the arrays of mean and standard deviation numbers: t_{dr}, t_v, t_{dl}, t.

$$t_{dr} = \Delta m_{dr}^2 + \Delta\sigma_{dr}^2 \quad t_v = \Delta m_v^2 + \Delta\sigma_v^2$$

$$t_{dl} = \Delta m_{dl}^2 + \Delta\sigma_{dl}^2 \quad t_h = \Delta m_h^2 + \Delta\sigma_h^2$$

   The variable m represents local gray level mean, and σ represents the local gray level standard deviation. The use of Δ indicates a local difference in the values of m or σ. The subscripts on m and σ (dr, v, dl, h) indicate the directions in which the differences were taken (diagonal to the left, vertical, diagonal to the right, and horizontal, respectively). To compute the differences, the spatial distance between values of m and σ is chosen as the window size used to compute m and σ (i.e., 7 pixels). The resulting value of texture gradient is the maximum of t_{dr}, t_v, t_{dl}, and t_h.
3. Generate an initial binary segmentation based upon two user-selected thresholds. All thresholds within the segmentation process are chosen by empirical means and take the form of percentages of the maximum of their respective variables.

4. Generate a thinned image by shrinking the binary segmentation image so that only a one-pixel wide skeleton of the segmentation remains.
5. Generate a multi-gray-level segmentation based on the evidences gathered by multiple (user-specified) thresholds applied to the mean image.
6. Combine the multi-gray-level segmentation with the skeletonized binary segmentation to create a constrained edge image.
7. Link the edges in the constrained edge image. The unconnected edge ends are identified, and for each edge end the closest unconnected end is found and the two are connected. The result is a map of the boundaries of the regions extracted in this phase of processing.

Region Merging

1. Create a new image by overlaying the region boundary image on the original image.
2. Identify all the bounded regions.
3. Describe the boundary of each region.
4. Merge all "small" regions with an adjacent region on the basis of the size of the common border. The small region size is defined empirically (usually chosen to be greater than 75 pixels).
5. Construct a new edge image to account for the merging process.
6. Compute the gray level mean and variance for each of the resulting regions.
7. Merge the regions based upon their computed features.

The result of the initial segmentation and region merging steps is a segmented image that will be used by the interest point selection algorithm.

2.1.2. *Interest Point Selection*

We compute a set of distinguishable points by passing an operator, which is a combination of the Hessian and Laplacian operators,[10] over each frame of imagery. The operator, I, takes the form

$$I(g) = g_{xy}^2 - g_{xx}g_{yy}$$

where g is the local gray level "function" and g_{xx}, e.g., is the local second derivative in the x direction. The interest operator, I, actually computes a measure of gray level curvature. In computing $I(g)$ for a particular image, the image is first smoothed by convolution with a small Gaussian kernel. The derivatives are then computed on the smoothed image by convolution by the 3×3 kernels

$$\Delta_{xx} = \frac{1}{3}\begin{bmatrix} 1 & -2 & 1 \\ 1 & -2 & 1 \\ 1 & -2 & 1 \end{bmatrix} \quad \Delta_{yy} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{bmatrix} \quad \Delta_{xy} = \frac{1}{2}\begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

This combination of smoothing and small kernel derivatives used in computing the interest operator has resulted in a more temporally consistent set of interest points being extracted than any other approach implemented by the authors. Stated in another way, the benefit of this particular implementation is that an interest point corresponding to a particular world feature will more consistently remain on that feature throughout multiple frames. Larger kernels were found to blur the imagery excessively and result in decreased resolution in interest point location.

The locations of interest points is determined by first locating all local maxima of $I(g)$. At each of the maxima with value greater than some threshold value, we search for the closest local minimum with a value approximately equal to the negative of the maximum value. The interest point lies where the line joining the maximum and minimum crosses $I(g) = 0$. At this approximate point of curvature inflection, an interest point is defined. Note that this point is not rounded to the location of the nearest pixel; the actual value in (F, y, z) space is recorded. Hence, we have sub-pixel accuracy in the computation of interest point location. The exact benefit of sub-pixel accuracy in interest point location has not been evaluated in light of image quantization error and image signal-to-noise ratio.

The interpolation between local maximum and minimum has the effect of smoothing the temporal interest point location variations. The temporal variations occur due to image quantization effects, sensor vibration, and motion distortion, to name a few. The smoothing enhances our ability to track interest points through multiple frames, and it improves the accuracy of the computed range values.

Our implementation of the I operator ranks the detected interest points by the magnitude of their corresponding local maximum. By this ranking, only the strongest n_j interest points within a particular image segment, j, are used to satisfy that segment's interest point allotment (see Section 2.1.1).

2.2. Point Feature Matching

Given a set of point features (i.e., interest points) within each image in a sequence of imagery, and the associated attitude and position at which each image was obtained, an algorithm has been developed that can robustly match the point features contained in consecutive frames and can track the matched features through multiple frames. In the following subsections, a description of the manner in which point features are processed prior to use by the feature matcher is provided along with a description of the matching process.

2.2.1. Interest Point Derotation

To aid the process of interest point matching, we must make it seem as though image plane $m + 1$ is parallel to image plane m. If this is done, the FOE and each pair of matched interest points between frames m and $m + 1$ would ideally be co-linear should the image planes be superimposed. Inertial data make this process possible.

The pixel
frame by th
the image p
that corresp
derotation t

$$\begin{bmatrix} F \\ y_j' \\ z_j' \\ 1 \end{bmatrix} =$$

where $(\phi, \theta$
and $m + 1$

$R,$

The matri
ments of
down (NI
The ma
compute
points de
within a c
frame m v
points in
into desc

2.2.2. In

The ma
first pass
interest
multiple
Hence, t
points in

The pixels in the image plane can be described in the sensor's 3-D coordinate frame by the vector (F, y, z), where F is the focal length of the sensor. To make the image planes parallel, derotation is performed for each vector, (F, y_j, z_j) that corresponds to each interest point in frame $m + 1$. The equation for the derotation transformation and projection (in homogeneous coordinates) is

$$\begin{bmatrix} F \\ y_j' \\ z_j' \\ 1 \end{bmatrix} = P\, R_{\phi_m} R_{\theta_m} R_{\psi_m} R^{-1}_{\psi_{m-1}} R^{-1}_{\theta_{m-1}} R^{-1}_{\phi_{m-1}} \begin{bmatrix} F \\ y_j \\ z_j \\ 1 \end{bmatrix} = P\, C^{m}_{NED}\, C^{NED}_{m+1} \begin{bmatrix} F \\ y_j \\ z_j \\ 1 \end{bmatrix}$$

where (ϕ, θ, ψ) are the roll, pitch, and yaw angles, respectively, of the frames m and $m + 1$, and where

$$R_\phi = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & \sin\phi & 0 \\ 0 & -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad R_\theta = \begin{bmatrix} \cos\theta & 0 & -\sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_\psi = \begin{bmatrix} \cos\psi & \sin\psi & 0 & 0 \\ -\sin\psi & \cos\psi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ F^{-1} & 0 & 0 & 0 \end{bmatrix}$$

The matrix P is the perspective projection transformation. All inertial measurements of sensor attitude rate and sensor velocity are made in a north–east–down (NED) coordinate frame.

The matrix P projects a world point onto an image plane and is used to compute the FOE, FOE $= P\,\vec{d}$, where $\vec{d} = \vec{v}\Delta t$. The matrix C^{m}_{NED} converts points described in the NED coordinate frame into an equivalent description within a coordinate frame parallel to the sensor coordinate frame when image frame m was acquired. Likewise, the matrix C^{NED}_{m+1} converts the descriptions of points in the sensor coordinate frame that corresponds to image frame $m + 1$ into descriptions in a coordinate frame parallel to the NED frame.

2.2.2. *Interest Point Matching*

The matching of interest points is performed in two passes. The goal of the first pass is to identify and store the top three candidate matches for each interest point in frame $n(= m + 1)$, (F, y_{n_i}, z_{n_i}). The second pass looks for multiple interest points, (F, y_{n_i}, z_{n_i}), being matched to a single point in frame m. Hence, the result of the second pass is a *one-to-one* match between the interest points in the two successive frames. For our application, a one-to-one match of interest points is necessary. We acknowledge that the projection of a world object onto the image plane of a sensor will grow in size as the sensor moves toward the object. This situation might imply that a one-to-one match does not make sense because what was one pixel in size in frame m might become two or more pixels in size in frame n. In this work, we assume that the growth of an object's projection, in terms of pixel count, is negligible. This assumption is based on the fact that our approach is point based, i.e., only one point (of sufficient interest) in each of two consecutive image frames is all that is required to fall upon an object's surface for the range to the object to be computed. The selection of the interest points is independent of object size.

The goal of interest point matching is to identify and store the best match in frame m for each interest point in frame n, (F, y_{n_j}, z_{n_j}). Several metrics/constraints assist us in this task. To determine the candidate matches to (F, y_{n_j}, z_{n_j}), each of the interest points in frame m is examined with the successive use of four metrics.

The *first metric* makes certain that candidate matches lie within a cone-shaped region, with apex at the FOE, bisected by the line joining the FOE and the interest point in frame n. The *second metric* constrains the distance between an interest point and its candidate matches. This is done by imposing maximum and minimum range constraints upon the resulting match.

The *third metric* requires that the interestingness, edge magnitude, and edge direction of both points of a candidate match are nearly equivalent. Edge direction is treated differently than the other parameters. We recognize that when an edge's normal is perpendicular to the line connecting the edge's pixels to the FOE, any interest points on this edge will not be reliably matched and ranged. This is due to the way interest points travel radially away from the FOE. The interest points along these special edges are weighted differently because they are more difficult to track and therefore less reliable.

The *fourth metric* restricts all candidate matches in frame m to lie closer to the FOE than the points in frame $n(= m + 1)$ (as physical laws would predict for stationary objects).

Hence, the first, second, and fourth metrics combine such that for an interest point in frame m, m_i, to be a candidate match to point n_j, m_i must lie in a region shaped like a sector of an annulus.

The reasoning behind the maximum and minimum range restriction is that world objects of range less than $R_{\min}$ are not possible considering the sensor mounting location on the vehicle and its field of regard. Stated another way, world objects that would lie closer than some $R_{\min}$ have been visible for some time and have been detected and therefore avoided by the vehicle navigator (machine or human). Likewise, objects at a range greater than $R_{\max}$ are not yet of concern to the vehicle.

2.2.3. Matching and Range Confidence Factors

We further improve range computations (based upon three or more sequential frames) by predicting and smoothing the range to each interest point that can be tracked through multiple frames. The procedure for prediction and smoothing of range using multiple frames is to compute, for all interest points in a pair of images, the matching confidence, confidence in range, and predicted ranges. Once the confidences and predicted range are computed, thresholds are applied and a smoothed range is computed.

The *matching confidence* of the ith point in frame m is given by

$$C_{Mi}^{m} = w_1\left[1 - \frac{|I_{mi} - I_{nj}|}{\max I_{mn} - \min I_{mn}}\right] + w_2\left[1 - \frac{|d_i^m - \min\limits_i d_i^m|}{\max\limits_i d_i^m - \min\limits_i d_i^m}\right] + w_3|\hat{o} \cdot \hat{a}|$$

where

$$\max I_{mn} = \max_i\,(I_{mi}, I_{ni}),\ \min I_{mn} = \min_i\,(I_{mi}, I_{ni}),$$

$$w_1, w_2, w_3 \geq 0 \text{ and } w_1 + w_2 + w_3 = 1$$

The variable I_{Xi} is the *interestingness* of the ith point in frame X and d_i is the perpendicular distance between the ith point and the line that passes through its matched point and the FOE. The unit vector $\hat{o}$ is in the direction of the line connecting the FOE and the ith point in frame m. The unit vector $\hat{a}$ represents the normal to the edge on which the ith interest point is located. The purpose of $|\hat{o} \cdot \hat{a}|$ is to cause the match confidence to fall when $\hat{o}$ and $\hat{a}$ are perpendicular (see Section 2.2.2).

The *range confidence*, C_{Ri}^{X}, of the ith point in frame X is given by the following set of equations

$$R_{i\,final}^{0} = R_{i\,predicted}^{0} = R_{i\,measured}^{0} \text{ and } C_{Ri}^{0} = 1 \tag{1}$$

$$R_{i\,predicted}^{n} = R_{i\,final}^{n-1} - velocity_i \times time \tag{2}$$

If $(R_{i\,predicted}^{n} \leq 0)$ then

$$R_{i\,final}^{n} = R_{i\,predicted}^{n} = R_{i\,measured}^{n} \text{ and } C_{Ri}^{n} = 0.5 \tag{3}$$

Else, if $(1 - \alpha < R_{i\,predicted}^{n}/R_{i\,measured}^{n} < 1 + \alpha)$ then

$$C_{Ri}^{n} = \frac{3}{2}\,C_{Mi}^{n}\left[C_{Ri}^{n-1} + \frac{1}{2}\left(1 - 2\left|\frac{R_{i\,measured}^{n} - R_{i\,predicted}^{n}}{R_{i\,measured}^{n} + R_{i\,predicted}^{n}}\right|\right)\right] \tag{4}$$

$$R_{i\,final}^{n} = R_{i\,measured}^{n} + (1 - C_{Ri}^{n})(R_{i\,predicted}^{n} - R_{i\,measured}^{n}) \tag{5}$$

If $(R^n_{i\,final} < 0)$ then

$$R^n_{i\,final} = \frac{R^n_{i\,measured} R^n_{i\,predicted}}{R^n_{i\,measured} + R^n_{i\,predicted}} \tag{6}$$

The variable α is a user-defined parameter that controls the range of the ratio $R_{i\,predicted}/R_{i\,measured}$.

2.3. Range Calculation and Interpolation

After the interest point matching process is complete, the matched pairs of interest points can be used to compute the range to the corresponding world objects. Given this collection of sparse range measurements, a range or obstacle map can be constructed. The obstacle map can take many forms,[11,12] the simplest of which consists of a display of bearing vs. range. In what follows, the range calculation is described and the important issue of range interpolation is discussed.

2.3.1. Range Calculation

Given pairs of interest points matches between two successive image frames and the translational velocity of the sensor during the time interval between frame acquisitions, it becomes possible to compute the range to the objects that correspond to the interest points. Our approach to range computation is described by the equation

$$R = \Delta X \frac{y' - y_f}{y' - y} \frac{1}{\cos \alpha_m} \tag{7}$$

where

y_f = the distance between the FOE and the center of the image plane,
y = the distance between the pixel in frame m and the center of the image plane,
y' = the distance between the pixel in frame $m + 1$ and the center of the image plane,
$\Delta X = |\vec{v}|\Delta t \cos \alpha_F$ = the distance traversed in one frame time, Δt, as measured along the axis of the line of sight,
α_F = the angle between the velocity vector and the line of sight,
α_m = the angle between the vector pointing to the world object and the line of sight,
$y' - y_f$ = the distance in the image plane between (F, y_{n_i}, z_{n_i}) and the FOE, and
$y' - y$ = the distance in the image plane between (F, y_{n_i}, z_{n_i}) and (F, y_{m_i}, z_{m_i}),

These variables are illustrated in Figure 4. This range equation is used to compute the distance to a world point relative to the lens center of frame m (a

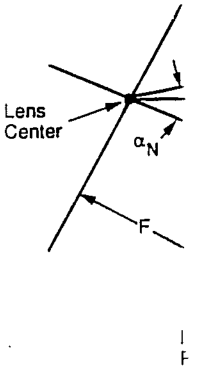


Figure 4. Geo
shows the ima
geometry.

similar equati
The accuracy
sensitive to th
process, the a

2.3.2. Range

The task o
passive rangi
be required b
systems). Th
between the
measurement
view. Essenti
of data point
the scene wit
samples be a
will require
understandin
computed m
The type o

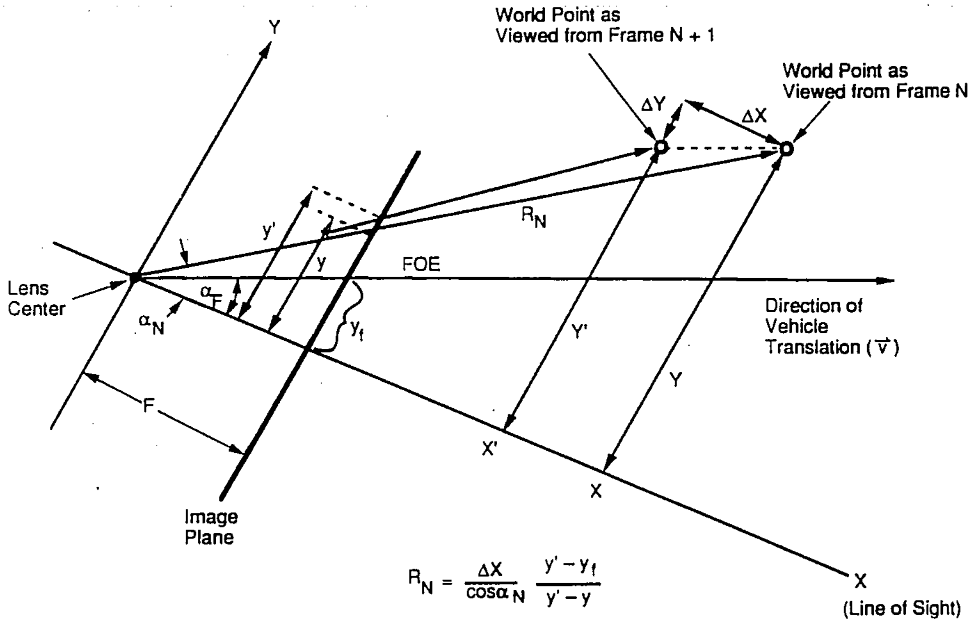


Figure 4. Geometry involved in the range calculation is illustrated here. The figure shows the imaged world point in motion rather than the sensor, thus simplifying the geometry.

similar equation would compute the distance from the lens center of frame n). The accuracy of the range measurements that are produced by eq. (7) is very sensitive to the accuracy of the interest point extraction process, the matching process, the accuracy of the INS data, and the accuracy of the sensor model.

2.3.2. Range Interpolation

The task of range interpolation is the last processing step required of the passive ranging system (this ignores any postprocessing of the range that may be required before it gets passed to the automatic vehicle control and display systems). The purpose of this task is to create, by means of interpolation between the sparse range samples generated from the motion analysis-based measurements, a dense range map representing the objects within the field of view. Essentially, this task is one of surface fitting to a sparse, nonuniform set of data points. To obtain an accurate surface fit that physically corresponds to the scene within the field of view, it is necessary that the sparse set of range samples be as uniformly spread throughout the field of view as possible. This will require processing steps described in previous sections, namely, scene understanding/segmentation must be used to create regions from which only a computed maximum number of interest points can be extracted.

The type of surface fitting is important because the resulting surface (i.e., the range map) must pass through each of the range samples. It would be especially dangerous (in the sense of obstacle avoidance) if the surface passed *under* any range samples. There are many techniques of surface fitting available to our task. To date, we have explored a method of bivariate interpolation over irregularly spaced samples proposed by Akimo.[13] This technique uses fifth degree polynomials to interpolate over the triangular regions formed by triangulation of the range sample locations. The major drawback associated with this approach is its assumption that all of the given points fall within a convex region. A solution to this problem is to use an improved Delaunay-based triangularization of the range samples, proposed by DeFloriani *et al.*,[14] that works over arbitrarily shaped regions of interest. Neither approach generates a surface that is guaranteed to pass through the range samples. The resulting surface is also quite undulating artificially due to the fifth degree polynomials that are used.

A less elaborate technique of range interpolation consists of fitting planar patches to the available range samples after performing a Delaunay triangulation of the samples. This approach gets the job done quickly and efficiently and does succeed in passing through each range sample, although the resulting surface, due to its planar patch construction, contains discontinuities in range.

All techniques of range interpolation should be careful to avoid interpolation over depth/range discontinuities that occur between range samples on the surface under investigation. With the use of scene analysis/segmentation, the smoothing of discontinuities can be avoided by interpolating only internal to the *smooth* regions or segments of the scene. Techniques of joining the regions after interpolation would then be employed. Such techniques have yet to be developed.

Finally, there is some concern as to the purpose of interpolation. Surely, interpolation will aid an operator/pilot in the interpretation of the results of passive range measurements, but its use by automatic vehicle control is in question. Also, a large number of interest points can be selected and matched, so there may not be any need for elaborate interpolation. These issues are being explored further.

3. EXPERIMENTAL RESULTS

Our inertial navigation sensor integrated motion analysis algorithm has been used to generate range samples from both indoor/laboratory imagery with simulated INS data and outdoor imagery with real INS data that was obtained from onboard a moving vehicle. In this section, we describe the conditions under which the data was created/collected and provide images illustrating the results of the major steps in the motion analysis algorithm.

3.1. Indoor Data

A sequence of imagery was collected inside of a computer lab by moving a camera forward in discrete 2.0-ft. steps. The velocity and attitude of the camera were estimated as 2 ft./s forward with no attitude changes throughout all five

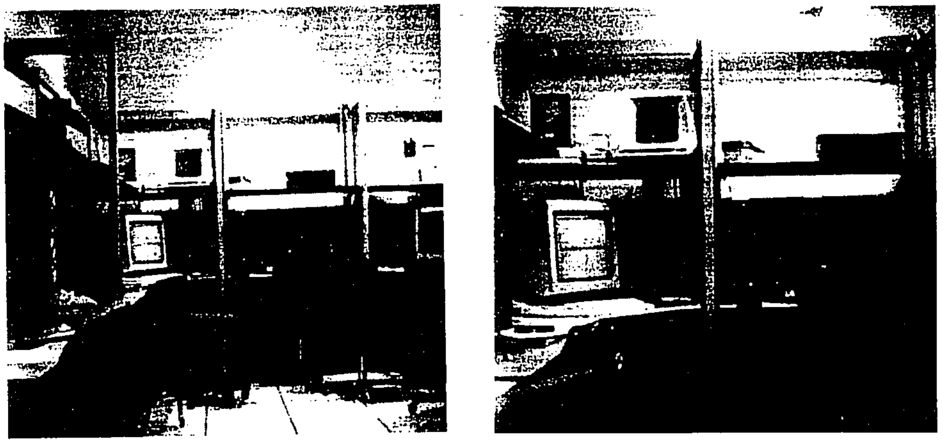


Figure 5. Five-frame sequence of indoor/lab imagery.

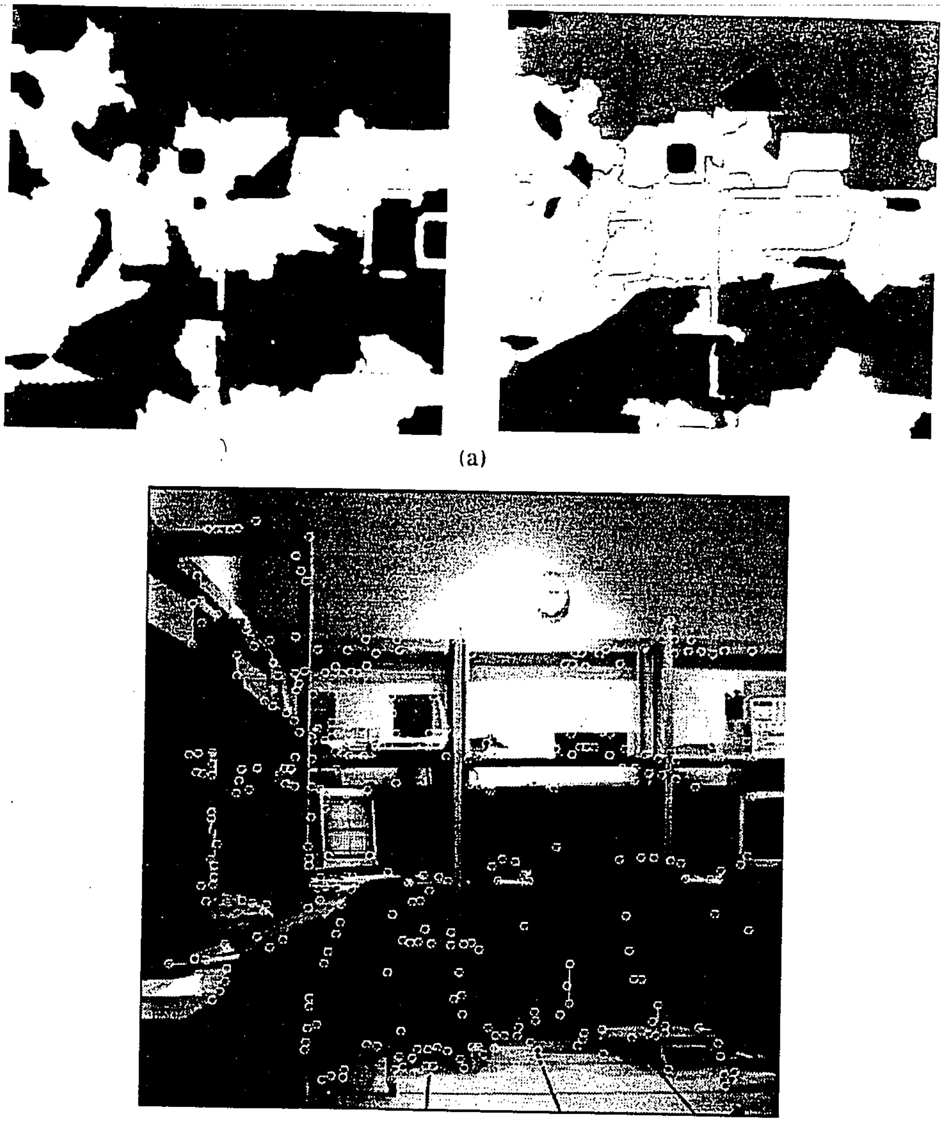
frames. The first and last frames of the resulting imagery are displayed in Figure 5. The field of view of the camera used to collect these images is 43.2° × 18° and the focal length is 12.5 mm. An example of the processing that was performed is displayed in Figure 6. The results of the various steps are illustrated: (a) segmentation, (b, c) interest point extraction, and (d) matching.

The image in Figure 7 is the cumulative result of processing the five-frame sequence. Ideally, we would see a chain of connected circles that would denote the location of strong interest points that were tracked through all five frames. In this case, we see very few instances of chains of circles due in part to the large separation between frames.

For the lab images, we have a limited amount of *ground truth* information. By ground truth, we mean that we have actually measured the range between the camera and various lab objects. With this information, we can begin to study the accuracy of the motion analysis-generated range values. Figure 8 shows the locations of the objects for which ground truth exists. Table I provides a comparison of ground truth range values and the range values generated through motion analysis for four pairs of imagery. Note that some motion analysis range values are missing. This is because a computer-selected interest point did not fall on the corresponding ground truthed object. For the table entries provided, there exists an interest point that fell on the corresponding ground truthed object.

Table II illustrates the effect that the smoothing filter has on interest point range values when a world object is tracked through multiple frames. The *final* range values, as described in Eqs. (1)–(6), are dependent upon the listed values of match confidence, range confidence, predicted range, and measured range.

The indoor/lab sequence of imagery, as described above, was not collected with a fixed configuration of hardware (i.e., INS and camera). Due to the large amount of movement of the sole camera, the accuracy of the "inertial data"


(a)

(b)

Figure 6. Results of processing one pair of the indoor imagery: (a) the segmentation of both frames, (b) the interest points in the first frame, (c) the interest points in the second frame, and (d) the set of matched points.

and the alignment of the camera between image acqusitions is largely responsible for the error between the ODIN-generated range measurements and the ground truth as shown in Table I. It is also important to note that a certain amount of error exists in the ground truth measurements because they were manually obtained and registered with the camera's coordinate frame. In gen-

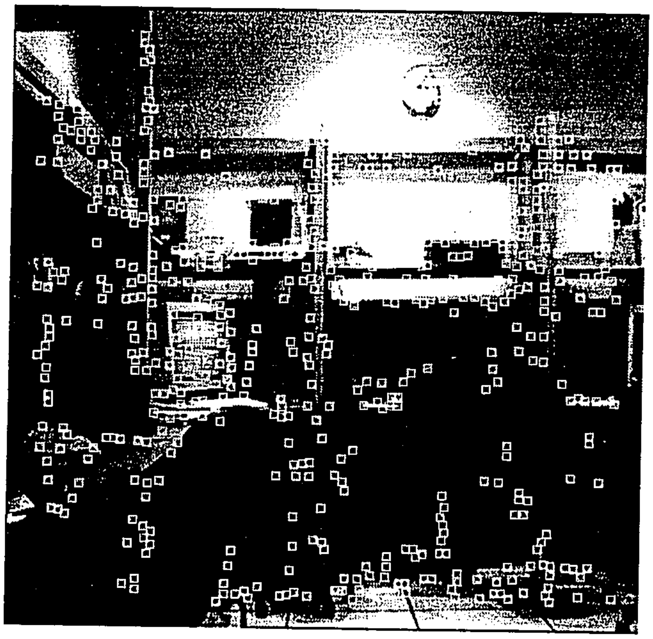


Figure 6. (*continued*)(c).

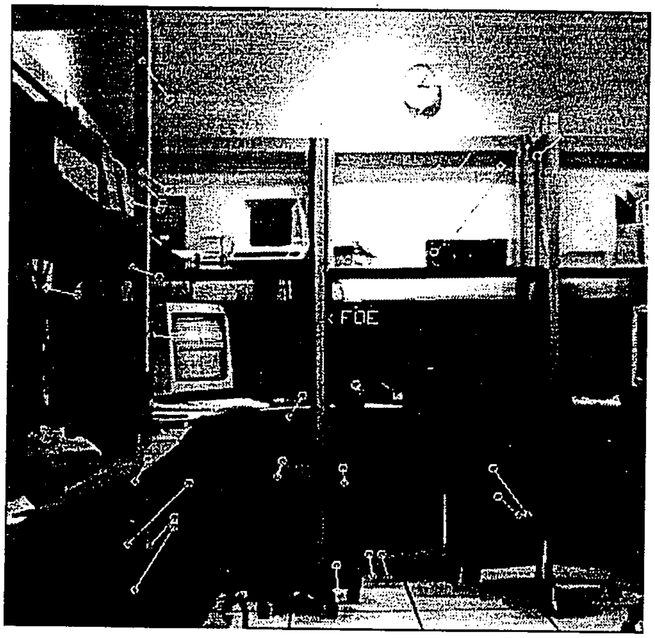


Figure 6. (*continued*)(d).

Table I. Comparison of ground truth and motion analysis range values for the indoor imagery.

Ground truth location	1–2 Range (ft.)		2–3 Range (ft.)		3–4 Range (ft.)		4–5 Range (ft.)	
	Actual	ODIN	Actual	ODIN	Actual	ODIN	Actual	ODIN
A	13.76	13.96	11.80	—	9.84	10.47	7.57	10.90
B	14.28	14.15	12.33	—	10.40	10.63	8.50	—
C	14.00	14.60	12.05	—	10.12	—	8.22	—
D	20.95	20.52	19.00	—	17.02	21.05	15.07	13.47
E	18.13	—	16.16	21.41	—	17.63	12.24	—
F	20.64	20.97	18.65	16.70	16.67	21.51	14.70	15.64
G	21.14	—	19.14	17.26	17.15	23.87	15.16	—
H	20.14	18.50	18.14	15.31	16.14	12.37	14.14	—
I	22.37	20.68	20.37	—	18.38	19.82	16.38	15.80
J	23.00	21.16	21.02	—	19.04	—	17.08	—
K	20.66	20.59	18.73	16.98	16.80	14.83	14.91	—

Columns labeled *Actual* contain the ground truth values and the columns labeled *ODIN* contain the motion analysis-generated range.

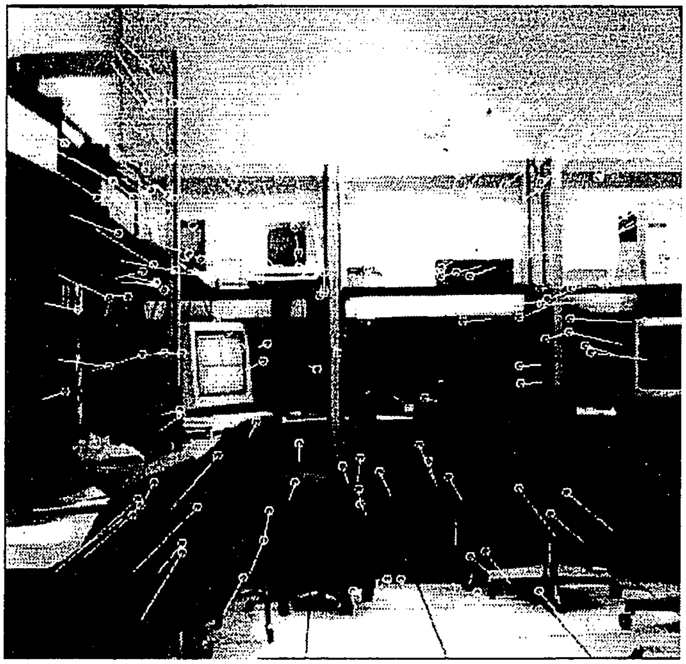


Figure 7. Cumulative result of processing five frames of indoor imagery. Every interest point that was matched and assigned a range is superimposed here on the first frame of the sequence.

Table II. Set of three world points that appear as interest points in three or more consecutive frames.

	Range (ft.)			
Range smoothing results	1–2	2–3	3–4	4–5
Measured	—	14.12	15.89	18.14
Predicted	—	—	12.17	13.22
Final	—	14.12	15.16	16.53
Match confidence	—	0.60	0.88	0.88
Range confidence	—	0.5	0.80	0.67
Measured	—	14.12	15.89	—
Predicted	—	—	12.17	—
Final	—	14.12	15.16	—
Match confidence	—	0.60	0.88	—
Range confidence	—	0.5	0.80	—
Measured	26.78	20.02	—	—
Predicted	—	24.80	—	—
Final	26.78	23.06	—	—
Match confidence	0.88	0.61	—	—
Range confidence	0.5	0.36	—	—

One can see the effect of the smoothing filter in generating the final range values and the filter's dependence upon the confidence factors.

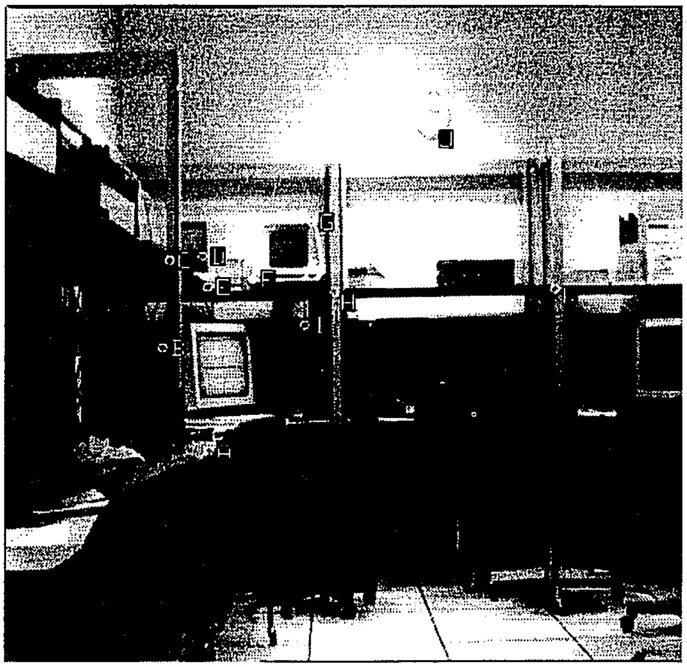


Figure 8. Locations of the lab points that have associated ground truth information. Video frames with time stamp are recorded at 5 Hz synchronously with IRU data. IRU data with time stamp is collected at 50 Hz.

Table III. Actual attitude and velocity measurements that were made in synchronism with the acquisition of five frames of outdoor imagery.

	Attitude (rad)			Velocity (ft./s)		
Frame	Roll	Pitch	Yaw	v_{north}	v_{east}	v_{down}
1	3.49e-02	2.72e-02	1.33	2.24	8.36	−0.149
2	2.99e-02	2.75e-02	1.328	2.30	8.32	−0.150
3	2.61e-02	2.90e-02	1.327	2.23	8.23	−0.150
4	2.42e-02	3.01e-02	1.326	2.19	8.23	−0.120
5	2.53e-02	2.99e-02	1.325	2.01	8.23	−0.133

These measurements are in the NED coordinate frame of the INS that shows that vehicle motion is roughly in the E–NE direction.

eral, we claim that the accuracy of the motion analysis-generated range measurements (obtained on the indoor imagery) is within 15% of the ground truth values, but the ground truth itself is subject to at least a 5% error.

3.2. Outdoor Data

A sequence of outdoor imagery was collected along with INS data generated by a Honeywell HG1050 inertial measurement unit. Table III indicates the roll, pitch, yaw, and velocity of the camera associated with the sequence of outdoor frames that were used. The velocity and attitude measurements are made in the coordinate frame of the INS. Figure 9 illustrates the hardware used to collect

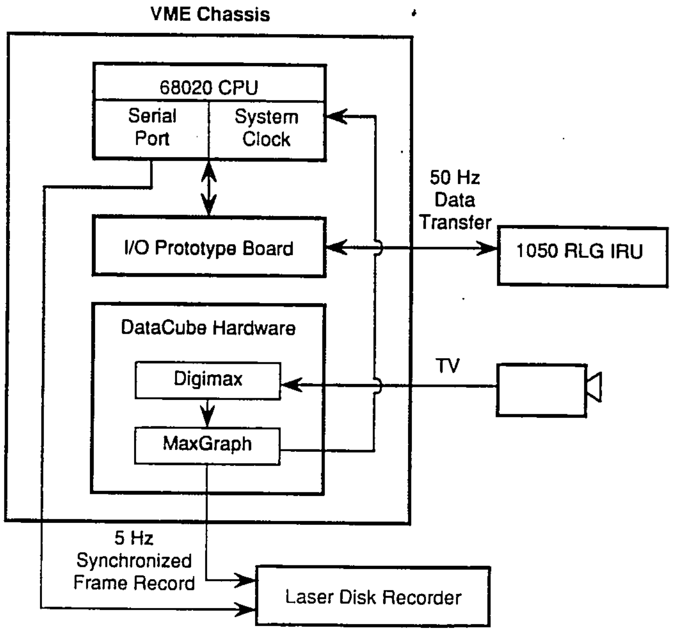


Figure 9. Hardware used to collect the outdoor imagery and INS data.

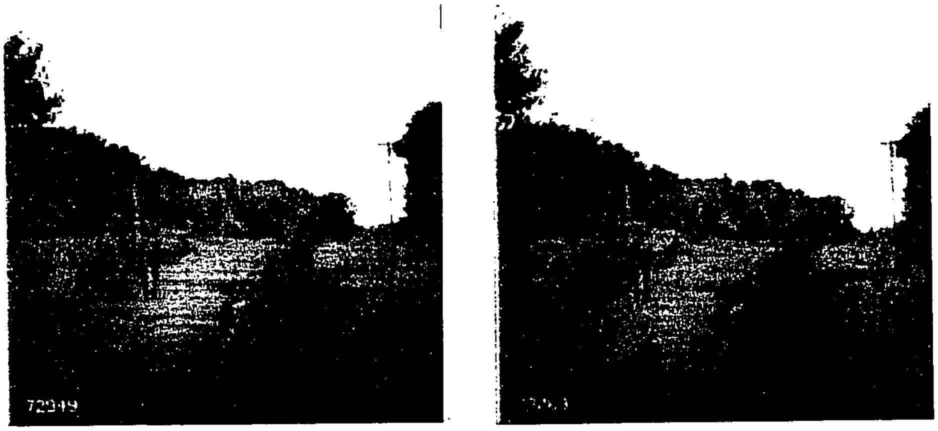


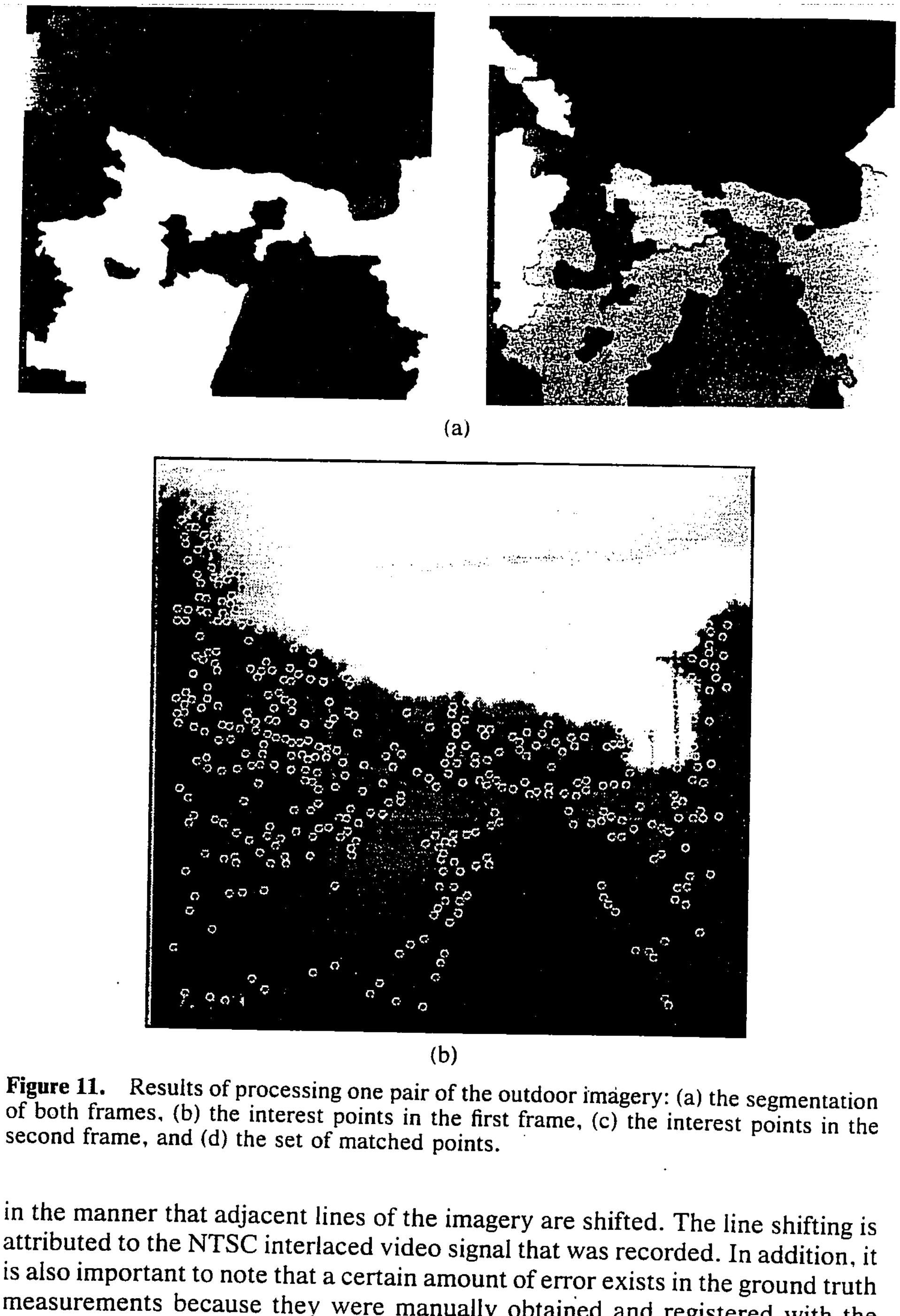
Figure 10. First and last frames in a five-frame sequence of outdoor imagery.

the imagery. The video frames were stored on optical disk at a 5-Hz rate that was synchronized with the collection of INS data. The INS data was collected at a 50-Hz rate and stored with a time stamp. To determine the correspondence between video frame and INS data packet, it is a simple matter to read the time stamp that was written on the frame when it was stored and then locate the corresponding INS data packet.

The first and last frames of a five-frame sequence of the collected imagery are presented in Figure 10. The field of view of the camera used to collect these images is 32.6° × 22.1° and the focal length is 15.1 mm. The elapsed time between each pair of frames for this experiment was 0.3 s. An example of the processing that was performed is displayed in Figure 11. The results of the various steps are illustrated: (a) segmentation, (b, c) interest point extraction and derotation, (d) matching, and (e) computed range. Note that only the interest points in the second frame of the pair are derotated. The derotated locations of the points are represented by diamonds and their original positions are shown as squares. The points in the first image of the pair are denoted by circles. The image in Figure 12 is the cumulative result of processing the five-frame sequence.

For the outdoor images in Figure 11, we also have a limited amount of *ground truth* information/data. These data were collected using a theodolite and were manually registered with the camera's coordinate frame. Figure 13 shows the locations of the objects for which ground truth exists. Table IV provides a comparison of ground truth range values and the range values generated through motion analysis for four pairs of imagery. Again, note that some motion analysis range values are missing because no interest points fell on the appropriate ground truthed object.

The outdoor sequence of imagery suffers from distortion due to camera motion and vibration. The distortion is visualized in the form of image blur and



(a)

(b)

Figure 11. Results of processing one pair of the outdoor imagery: (a) the segmentation of both frames, (b) the interest points in the first frame, (c) the interest points in the second frame, and (d) the set of matched points.

in the manner that adjacent lines of the imagery are shifted. The line shifting is attributed to the NTSC interlaced video signal that was recorded. In addition, it is also important to note that a certain amount of error exists in the ground truth measurements because they were manually obtained and registered with the camera's coordinate frame. These combined effects are largely responsible for the error between the ODIN-generated range measurements and the ground

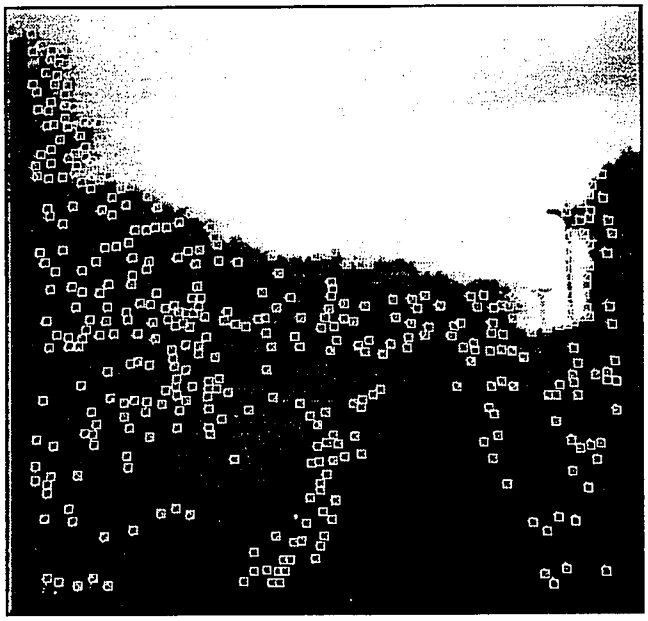


Figure 11. (*continued*)(c).

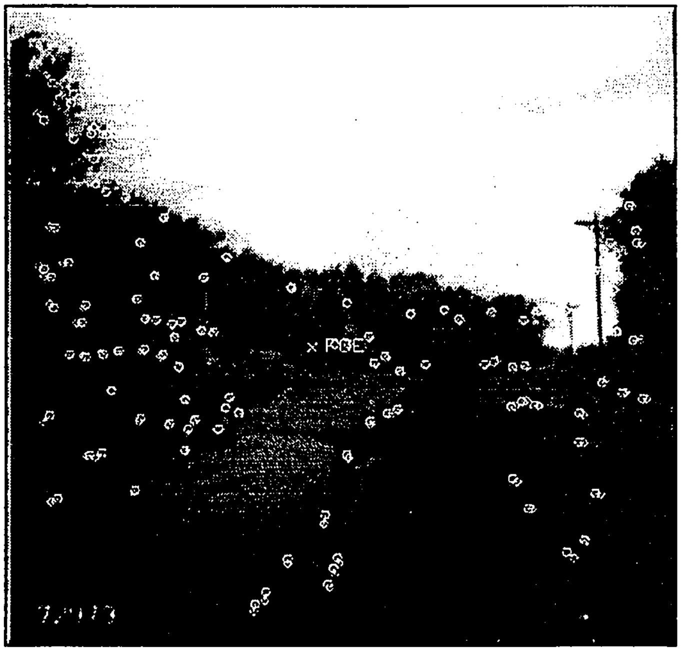


Figure 11. (*continued*)(d).

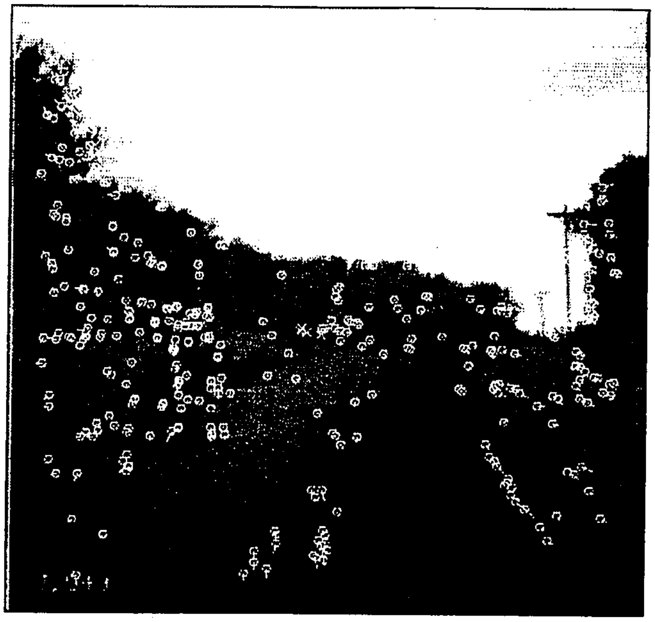


Figure 12. Cumulative result of processing five frames of outdoor imagery. Every interest point that was matched and assigned a range is superimposed here on the first frame of the sequence.

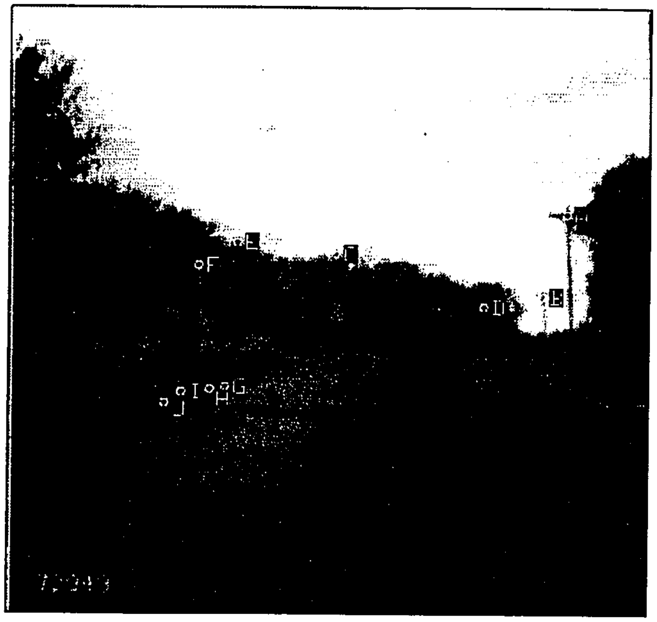


Figure 13. Locations of the world points that had associated ground truth information.

Table IV. Co

Ground truth location
A. Telephone pole
B. Telephone pole
C. Treeline #
D. Treeline #
E. Treeline #
F. Pole by ga
G. Red light post (closes to road)
H. Red light post (closes to gate)
I. Fence post gate (west end, closes
J. Fence pos gate (west end, farthe

The column contain the mo

truth as sh
motion ana
ery) is with
subject to

4. CONCL

In this a
presented.
incorporati
the analys
scene ana
interpolati
of the mot
tion and to
Our ong
system for
commerci
be perfor
suite. In s

Table IV. Comparison of ground truth and motion analysis range values.

Ground truth location	1–2 Range (ft.)		2–3 Range (ft.)		3–4 Range (ft.)		4–5 Range (ft.)	
	Actual	ODIN	Actual	ODIN	Actual	ODIN	Actual	ODIN
A. Telephone pole	231	185	228	276	226	245	223	202
B. Telephone pole	486	367	483	366	480	427	478	—
C. Treeline #1	502	—	499	—	497	430	494	—
D. Treeline #2	665	300	663	298	660	—	658	446
E. Treeline #4	388	—	385	—	383	255	380	—
F. Pole by gate	214	298	212	—	209	236	206	175
G. Red light post (closest to road)	153	186	150	322	148	—	145	165
H. Red light post (closest to gate)	156	167	153	343	151	114	148	157
I. Fence post by gate (west end, closest)	169	160	167	172	164	—	162	161
J. Fence post by gate (west end, farthest)	156	209	153	—	151	165	148	153

The columns labeled *Actual* contain the ground truth values and the columns labeled *ODIN* contain the motion analysis-generated range.

truth as shown in Table IV. In general, we claim that the accuracy of the motion analysis-generated range measurements (obtained on the outdoor imagery) is within 25% of the ground truth values, but the ground truth itself is subject to at least a 5% error.

4. CONCLUSIONS

In this article, our latest work on INS integrated motion analysis has been presented. The most important lesson learned from this research is that the incorporation of inertial data into the motion analysis problem greatly improves the analysis and makes the process more robust. In addition, the benefit of scene analysis as a tool to guide the interest point extraction and surface interpolation has been learned, and we have gained insight into the sensitivity of the motion analysis-based range computation to shifts in interest point position and to INS errors.

Our ongoing efforts at Honeywell include the development of a real-time system for passive ranging that is to be implemented in a single VME chassis of commercial off-the-shelf hardware. With the real-time system, flight testing can be performed to validate the performance of the passive ranging algorithm suite. In support of the validation process, additional data collection efforts are required to produce the high-quality data and the accompanying ground truth data needed for validation. Additional ongoing efforts include algorithm enhancements in the areas of feature selection, feature matching, and range interpolation.

This material is based upon work supported by NASA Ames Research Center under Contract NAS2-12800. The authors thank Dave Duncan, J. C. Ming, and Scott Snyder for their contributions to this work.

References

1. M. J. Daily, J. G. Harris, and K. Reiser, "Detecting obstacles in range imagery," in *Proc. of DARPA Image Understanding Workshop*, February 1987, pp. 87–97.
2. "Millimeter-wave radar may enhance safety of helicopter flights," *Aviation Week and Space Technology*, July 1987.
3. C. Thorpe, S. Schafer, and T. Kanade, "Vision and navigation for the Carnegie Mellon Navlab," in *Proc. of DARPA Image Understanding Workshop*, February 1987, pp. 143–153.
4. B. Bhanu and B. Roberts, "Obstacle detection during rotorcraft low-altitude flight," annual technical report to NASA-Ames, April 1989.
5. B. Bhanu, B. Roberts, and J. C. Ming, "Inertial navigation sensor integrated motion analysis," *Proc. of DARPA Image Understanding Workshop*, May 1989, pp. 747–763.
6. R. Dutta, R. Manmatha, E. M. Riseman, and M. A. Snyder, "Issues in extracting motion parameters and depth from approximate translational motion," in *Proc. of DARPA Image Understanding Workshop*, April 1988, pp. 945–960.
7. L. Matthies, R. Szeliski, and T. Kanade, "Kalman filter-based algorithms for estimating depth from image sequences," in *Proc. of DARPA Image Understanding Workshop*, April 1988, pp. 199–213.
8. J. A. Farrell, *Integrated Aircraft Navigation*, Academic Press, New York, 1976.
9. B. Bhanu and P. Symosek, "Interpolation of terrain using hierarchical symbolic grouping for multi-spectral images," in *Proc. of DARPA Image Understanding Workshop*, February 1987, pp. 466–474.
10. H. H. Nagel, "Displacement vectors derived from second-order intensity variations in image sequences," *Comp. Vision, Graph. Image Proc.* **21,** 85–117 (1983).
11. V. H. L. Cheng, "Obstacle-avoidance automatic guidance—a concept-development study," in *Proc. of AIAA Guidance, Navigation and Control Conf.*, August 1988, pp. 1142–1152.
12. F. W. Smith and M. Streicker, "Passive ranging from a moving vehicle via optical flow measurement," *SPIE: Appl. Digit. Image Proc.*, **829,** 310–317 (1987).
13. H. Akimo, "A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points," *ACM Trans. Math. Software*, **4,** 148–159 (1978).
14. L. DeFloriani, B. Falcidieno, and C. Pienovi, "Delaunay-based representation of surfaces defined over arbitrarily shaped domains," *Comp. Vision, Graph. Image Proc.*, **32,** 127–140 (1985).

Is Visual
Obstacle
Passive

Yiannis Aloim
Computer Visi
Computer Sci
Computer Stu
University of
College Park,

Received Feb

Is it possible t
necessary to cr
recognize, navi
without recons
collision in son
article, it is sh
detailed algori
tives of the im

受動レンジン
ル・システム
どの疑問に対
詳細について
回避のところ

1. INTROD

We are u
their visual
representati
much domi
preprocessi
of surfaces
tion about r
memory, e
needs. In t