INERTIAL NAVIGATION SENSOR INTEGRATED MOTION ANALYSIS FOR OBSTACLE DETECTION

Bir Bhanu, Barry Roberts, and John Ming

Honeywell Systems and Research Center
3660 Technology Drive
Minneapolis, MN 55418

ABSTRACT

Land navigation requires a vehicle to steer clear of trees, rocks, and man-made obstacles in the vehicle's path while vehicles in flight, such as helicopters, must avoid antennas, towers, poles, fences, tree branches, and wires strung across the flight path. Automatic detection of these obstacles and the necessary guidance and control actions triggered by such detection would facilitate autonomous vehicle navigation. An approach employing a passive sensor for mobility and navigation is generally preferred in practical applications of these robotic vehicles. Motion analysis of imagery obtained during vehicle travel can be used to generate range measurements, but these techniques are not robust and reliable enough to handle arbitrary image motion caused by vehicle movement. However, many types of existing vehicles contain inertial navigation systems (INS) which can be utilized to improve interest point selection, matching of the interest points, and the subsequent motion detection, tracking, and obstacle detection. We discuss an inertial sensor integrated optical flow technique for motion analysis to achieve increased effectiveness in obstacle detection during vehicle motion.

1. INTRODUCTION

A desired obstacle detection system for many practical applications should exhibit robustness and should not place unduly excessive size, power, or weight demands on the host vehicle. It should work in day/night/adverse weather conditions and should preferably be covert to minimize the threat to the vehicle and the pilot. The technique used for obstacle detection must also have graceful degradation, instead of total failure, under conditions of limited operability. In recent years, considerable effort has been put toward the detection of obstacles that present themselves primarily to ground vehicles. Using mainly active sensors, such as a laser scanner, obstacles (like fence posts, rocks, vegetation) are detected within the field of view of the vehicle's sensor.[3,8] Other active sensors such as Millimeter Wave (MMW) can detect obstacles such as wires, but the constant and continuous image of these active sensors may betray vehicle covertness.

Passive sensors, such as a TV camera, are also being used to detect obstacles for ground vehicles.[1,4,5] 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 place unrealistic constraints on the input data in order to make them work. The largest sources of error are sensor motion and incomplete/ambiguous information in the sensed image data. Many types of land and air vehicles (e.g. helicopters and military ground vehicles) contain an Inertial Navigation System (INS) whose output can be used for applications beyond the original intent of the system. Such vehicles can use the INS information to greatly simplify some of the functionalities normally provided by computer vision, such as obstacle detection, target motion detection, target tracking, stereo, etc. The INS used for the work reported here (a Honeywell HG1050) provides the vehicle's positional, velocity, and acceleration information, which enhances the quality of motion analysis techniques for obstacle detection.

The objective of the work presented in this paper is to describe our maximally passive approach to obstacle detection and to discuss the details of our inertial sensor integrated optical flow analysis technique.

2. INERTIAL SENSOR INTEGRATED MOTION ANALYSIS

The block diagram of the inertial sensor integrated motion analysis system is illustrated in Figure 1. The system uses inertial sensor integrated optical flow, scene analysis, and selective applications of active sensors to provide obstacle detection capability.[1] In this paper, we focus on the details of the inertial sensor integrated optical flow algorithm, which computes range to features within the sensor's field of view. For a pair of image frames, the major steps that are involved within the optical flow algorithm are given below:

(1) Input frames, frame A and frame B, are read in along with their associated inertial data.

(2) Interest points are extracted from each input frame.

(3) Location of the focus of expansion (FOE) (in both frames) is computed· based on the velocity vector obtained from the INS.

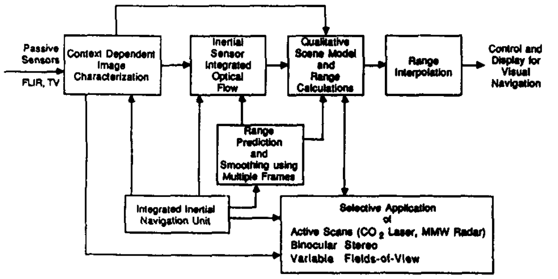

Passive Sensors
FLIR, TV
Context Dependent Image Characterization
Inertial Sensor Integrated Optical Flow
Qualitative Scene Model and Range Calculations
Range Interpolation
Control and Display for Visual Navigation
Range Prediction and Smoothing using Multiple Frames
Integrated Inertial Navigation Unit
Selective Application of
Active Scans (CO$_2$ Laser, MMW Radar)
Binocular Stereo
Variable Fields-of-View

Figure 1: Inertial sensor integrated optical flow and scene analysis provides robust image analysis useful for obstacle detection and avoidance.

CH2876-1/90/0000/0954$01.00 © 1990 IEEE

(4) The FOE and the interest points in frame B are projected onto an image plane that is parallel to the image plane that captured frame A (*derotation* of frame B).

(5) Interest points in frame B are matched to those of frame A based upon four criteria.

(6) Range is computed to each interest point in frame B that has a match in frame A.

(7) A dense range map is created using context dependent scene analysis and interpolation between the computed range values.

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 where 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 2.

As shown in Figure 3, 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 information, 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 described in the following section.

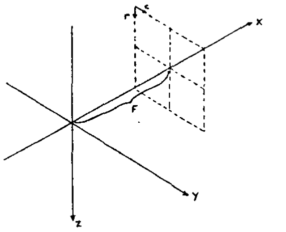


Figure 2: The coordinate system geometry of the sensor's image plane is perpendicular to the x axis, located at the distance of the focal length, F, from the origin.

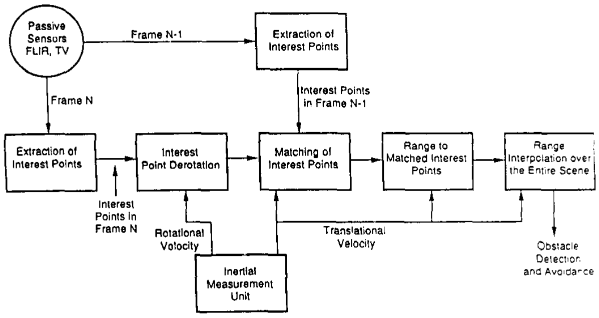

Passive Sensors FLIR, TV
Frame N-1
Extraction of Interest Points
Interest Points in Frame N-1
Frame N
Extraction of Interest Points
Interest Point Derotation
Matching of Interest Points
Range to Matched Interest Points
Range Interpolation over the Entire Scene
Interest Points in Frame N
Rotational Velocity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

Figure 3: Inertial sensor integrated optical flow technique.

2.1 DISTINGUISHED FEATURES

The features within the imagery (TV or FLIR) that are most prominent and distinguished, mark the world points to which range measurements will be made. These prominent world points, known as *interest points*, are easy to extract from the imagery and have the highest promise of repeated extraction throughout multiple frames. The interest points within the field-of-view of the monocular sensor are of fundamental and critical importance to optical flow calculations. In the following subsections, the extraction and subsequent use of interest points is described in detail.

2.1.1 Interest Point Selection - We compute a set of distinguishable points by passing an operator which is a combination of the Hessian and Laplacian operators[6] over each frame of imagery. The operator, I, takes the form

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

where g is the local gray level function and g_{xx} for example, is the local 2nd derivative in the x direction.

Our implementation of the interest point operator ranks the detected interest points in the order of their computed interestingness, S. This interest point extraction routine takes as input a segmentation of the original image and returns n_j, $0 \le j \le N$, interest points in each of the N segments. The value of n_j for segment j is proportional to the segment size and segment features. More than n_j interest points can exist per segment; only the points with the highest S values are reported. The result of returning only the best interest points (in terms of S) in each segment is that the processed scene is more uniformly covered with interest points.

Unfortunately, not all regions within a scene can contain reliable interest points. Scene analysis techniques are used to ascertain the *goodness* of regions prior to interest point selection.[1] Moreover, interest point selection is further improved by incorporation of Kalman filtering techniques, which use inertial sensor data to track and predict interesting point features.[1]

2.1.2 Interest Point Derotation - To aid the process of interest point matching, we must make it seem as though image plane B is parallel to image plane A (see Figure 4). If

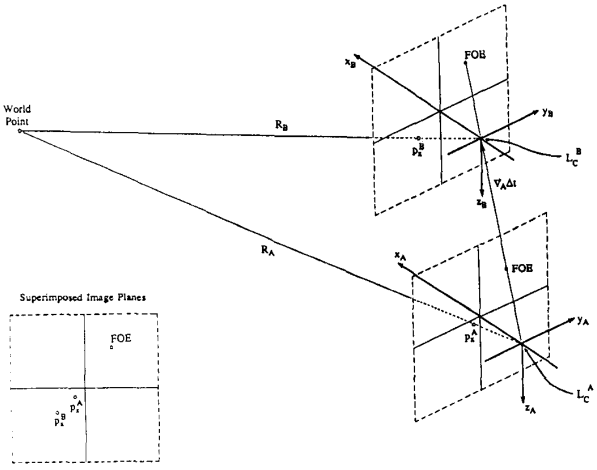


Figure 4: An illustration of the sensor geometry that records two perspective views of a scene at two positions separated by a distance $|\vec{v}|\Delta t = |\vec{d}|$ (with no rotation of the sensor between positions).

this is done, the FOE and pairs of interest points in frames A and B that match, would ideally be colinear should the image planes be superimposed. To make the image planes parallel, derotation is performed for each vector, (F, y_i, z_i) that corresponds to each interest point in frame B. The equation for the derotation transformation and projection (in homogeneous coordinates) is

$$\begin{bmatrix} F \\ y_i' \\ z_i' \\ 1 \end{bmatrix} = P\ R_{\phi_A}^{-1}\ R_{\theta_A}^{-1}\ R_{\psi_A}^{-1}\ R_{\psi_B}\ R_{\theta_B}\ R_{\phi_B} \begin{bmatrix} F \\ y_i \\ z_i \\ 1 \end{bmatrix}$$

$$= P\ C_{NED}^{A}\ C_{B}^{NED} \begin{bmatrix} F \\ y_i \\ z_i \\ 1 \end{bmatrix}$$

where

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

$$R_\psi = \begin{bmatrix} \cos\psi & -sin\,\psi & 0 & 0 \\ \sin\psi & \cos\psi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/F & 0 & 0 & 0 \end{bmatrix}$$

and where NED (*north, east, down*) is the coordinate frame in which inertial measurements are made. Use of the NED frame assumes that vehicle motion is "local" to a patch of Earth.

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_{NED}^{A} converts points described in the NED coordinate frame into an equivalent description within a coordinate frame parallel to the A coordinate frame. Likewise, the matrix C_B^{NED} converts the descriptions of points in the B coordinate frame into descriptions in a coordinate frame parallel to NED.

2.1.3 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 B, (F, y_{B_j}, z_{B_j}). The second pass looks for multiple interest points being matched to a single point in frame A. 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 onto the sensor's image plane of an object in the world 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 since what was one pixel in size in frame A might become two or more pixels in size in frame B. In this work, we assume that the growth of objects, in terms of pixel size, is negligible in the passive ranging for obstacle detection scenario. All objects are assumed to be at certain safe distances for vehicle maneuvering and one pixel (of interest point quality) in two frames is all that is required of an object's surface for the range to be computed.

Pass One: To determine the candidate matches to (F, y_{B_j}, z_{B_j}), each of the interest points in frame A is examined with the successive use of four metrics. The *first metric* makes certain that candidate matches lie within a cone shaped region bisected by the line joining the FOE and the interest point in frame B. This metric limits candidate matches to lie within the cone with apex at the FOE, as shown in Figure 5(a). If an interest point in frame A, (F, y_{A_i}, z_{A_i}), passes the first metric, then the second metric is applied to it. The *second metric* requires that the interestingness of candidate matches is close to the interestingness of the point that we are trying to match.

The *third metric* restricts all candidate matches in frame A to lie closer to the FOE than the points in frame B (as physical laws would predict for stationary objects). This metric can compute the distances of the interest points from the FOE in two different ways:

(1) The direct euclidean distance, d_1, between (F, y_{A_i}, z_{A_i}) and (F, y_{B_j}, z_{B_j}), and

(2) the distance d_2 which is the projection of d_1 onto the line joining (F, y_{B_j}, z_{B_j}) and the FOE.

The distance measures are graphically illustrated in Figure 5(b). Regardless of the way that the distance measure is computed, it can be used to identify the closest candidate matches to (F, y_{B_j}, z_{B_j}).

The *fourth metric* constrains the distance between an interest point and its candidate matches. For an interest point in frame A, A_j, to be a candidate match to point B_j, it must lie within the shaded region of Figure 5(a). The depth of the region is determined by this fourth metric while the width of the region is fixed by an earlier metric. By limiting interest points, A_j, to lie in the shaded region, we have effectively restricted the computed range of resulting matches to lie between $R_{\max}$ and $R_{\min}$. World objects at ranges less than $R_{\min}$ should not occur due to autonomous or manual navigation of the vehicle, thus avoiding potential collisions. Likewise, objects at a range greater than $R_{\max}$ are not yet of concern to the vehicle.

The result of the first pass of interest point matching is a list, for each (F, y_{B_j}, z_{B_j}), of three or fewer candidate matches that pass all metrics and have the smallest distance measures of all possible matches.

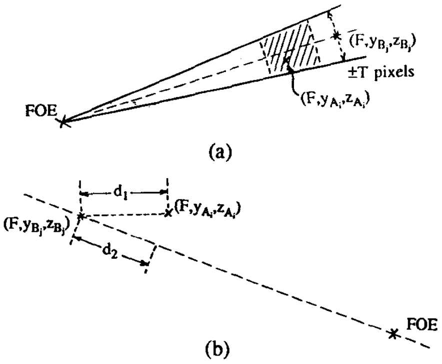


Figure 5: Constraints used to aid the process of matching interest points between frames. *(a)* Since an interest point, the FOE, and a candidate match must be colinear after derotation, all candidate matches to a point in frame B must lie within a cone with apex at the FOE and the shaded section. *(b)* There are two ways to compute the distance between interest points, distance metric d_1 or d_2.

Pass Two: The goal of the second pass of the matching process is to take the matches provided by the first pass and generate a *one-to-one* mapping between the interest points in frames A and B. Initially, it can be assumed that the best match to (F, y_{B_j}, z_{B_j}) will be the stored candidate match which has the smallest distance measure. Unfortunately, there may be multiple points, (F, y_{B_j}, z_{B_j}), which match to a single (F, y_{A_i}, z_{A_i}). Hence, the recorded list of best matches is searched for multiple occurrences of any of the interest points in frame A. If multiple interest points in frame B have the same best match, then the point, B^*, which is at the minimum distance from the A_i in question, will retain this match and is removed from the matching process. The remaining B_j's are returned to the matching process for further investigation after having A_i removed from their lists of best matches. This process continues until all of the interest points in frame B either have a match, or are determined to be unmatchable by virtue of an empty candidate match list. Hence, the final result of the matching process is a *one-to-one* mapping between the interest points in frames A and B.

2.2 RANGE CALCULATION AND INTERPOLATION

Given the result of interest point matching, which is the optical flow, range can be computed to each match. Given these sparse range measurements, a range or obstacle map can be constructed. The obstacle map can take many forms,[2,7] the simplest of which consists of a display of bearing versus range.

Given pairs of interest point matches between two successive image frames and the translational velocity between frames, it becomes possible to compute the range to the object on which the interest points lie. One approach to range, R, computation is described by the equation

$$R = \Delta\frac{Z}{\cos\alpha_A}\,\frac{x' - x_f}{x' - x}$$

where

x_f = the distance between the FOE and the center of the image plane,

x = the distance between the pixel in frame A and the center of the image plane,

x' = the distance between the pixel in frame B and the center of the image plane,

$\Delta Z = |\vec{v}|\Delta t \cos\alpha_F$ = 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,

$x' - x_f$ = the distance in the image plane between (F, y_{B_j}, z_{B_j}) and the FOE, and

$x' - x$ = the distance in the image plane between (F, y_{B_j}, z_{B_j}) and (F, y_{A_i}, z_{A_i}).

These variables are illustrated in Figure 6. Note that R is the distance to a world point relative to the lens center of frame A (similar equations would compute the distance from the lens center of frame B). The accuracy of the range measurements that result from either approach is very sensitive to the accuracy of the matching process as well as the accuracy of the inertial measurement unit (IMU) data.

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 optical flow 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 is a fundamental reason for our segmentation driven interest point selection.

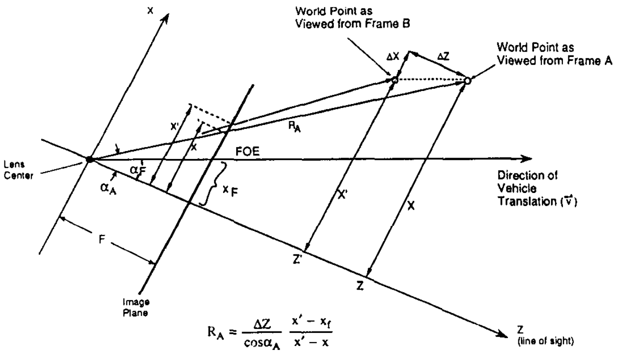

X
World Point as Viewed from Frame B
ΔX
ΔZ
World Point as Viewed from Frame A
x'
x
R_A
FOE
Lens Center
α_F
α_A
x_F
X'
X
Direction of Vehicle Translation (v⃗)
F
Z'
Z
Image Plane
R_A ≈ ΔZ/cosα_A · (x' − x_f)/(x' − x)
Z (line of sight)

Figure 6: The geometry involved in the first approach to range calculation, which shows the imaged world point in motion rather than the sensor, thus simplifying the geometry.

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 if the surface passed under any range samples. There are many techniques of surface fitting available to our task. However, there is some concern as to the purpose of interpolation. Surely, interpolation will aid an operator/pilot in the interpretation of the results of optical flow measurements, but its use by automatic vehicle control is questionable. Also, a large number of interest points can be selected and matched, so there may not be any need for interpolation. These issues are being explored further.

3. RESULTS

Our inertial navigation sensor integrated optical flow algorithm has been used to generate range samples using both synthetic data and real data (imagery and INS information) obtained from 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 optical flow algorithm.

The synthetic interest points were generated from a file containing the 3-D coordinates of 15 world points. Table 1 shows the 3-D locations of these world points. In the same coordinate system as the interest points are located, Table 2 lists the location, roll, pitch, and yaw of the camera at the two instances of time at which frames A and B were acquired. The time between frame acquisition is 0.2 seconds. Figure 7(a) shows the locations (circles) of the projection of the world points onto the first location of the image plane where the field of view of the synthesized camera model is 52.0° x 48.75° with a focal length of 9 mm. Figure 7(b) shows the locations (squares) of the projections of the world points onto the second location of the image plane and shows the new locations (diamonds) of those projections after derotation. Figure 7(c) shows the results of the matching process in which circles are connected to their corresponding diamond with a straight line and the FOE is labeled and marked with an X.

A pair of real images was selected to test the capabilities of the optical flow algorithm using real imagery. Table 3 indicates the location, roll, pitch, and yaw of the camera associated with the pair of real image frames that were used. The field of view of the camera for the real images is 52.1° x 40.3° and the focal length = 9 mm. The elapsed time between the two frames for this experiment was 0.2 seconds. Figure 8(a) shows the locations of the extracted interest points obtained from the first frame, drawn as circles. Similarly, Figure 8(b) indicates the location of extracted interest points (squares) and the corresponding derotated locations (diamonds). Since the vehicle undergoes very little rotation between frames, the derotated locations are nearly coincident with the original point locations. The results of the point matching process for the real imagery is shown in Figure 8(c).

4. CONCLUSIONS

We have presented out initial work for INS integrated motion analysis. Future work will involve incorporating context dependent qualitative scene analysis, knowledge-based sensor management, and incorporation of Kalman filtering into our approach. Our ultimate goal is to develop the complete, fieldable system for obstacle detection during rotorcraft low altitude flight. We are also applying this technology for land vehicle applications to achieve robust obstacle detection, target motion detection, and target tracking.

REFERENCES

1. B. Bhanu and B. Roberts, "Obstacle Detection During Rotorcraft Low-Altitude Flight," Annual Technical Report for NASA-Ames (April, 1989).
2. V.H.L. Cheng, "Obstacle-Avoidance Automatic Guidance - A Concept-Development Study," Proceedings of AIAA Guidance, Navigation and Control Conference, pp. 1142-1152 (August 1988).
3. M.J. Daily, J.G. Harris, and K. Reiser, "Detecting Obstacles in Range Imagery," Proceedings of DARPA Image Understanding Workshop, pp. 87-97 (February 1987).
4. R. Dutta, R. Manmatha, E.M. Riseman, and M.A. Snyder, "Issues in Extracting Motion Parameters and Depth from Approximate Translational Motion," Proceedings of DARPA Image Understanding Workshop, pp. 945-960 (April, 1988).
5. L. Matthies, R. Szeliski, and T. Kanade, "Kalman Filter-Based Algorithms for Estimating Depth from Image Sequences," Proceedings of DARPA Image Understanding Workshop, pp. 199-213 (April, 1988).
6. H.H. Nagel, "Displacement Vectors Derived from Second-Order Intensity Variations in Image Sequences," *Computer Vision, Graphics, and Image Processing* **21** pp. 85-117 (1983).
7. F.W. Smith and M. Streicker, "Passive Ranging from a Moving Vehicle via Optical Flow Measurement," *SPIE: Applications of Digital Image Processing* **829** pp. 310-317 (1987).
8. C. Thorpe, S. Schafer, and T. Kanade, "Vision and Navigation for the Carnegie Mellon Navlab," Proceedings of DARPA Image Understanding Workshop, pp. 143-153 (February 1987).

ACKNOWLEDGEMENTS

This material is based upon work supported by NASA under contract NAS2-12800 and DARPA/U.S. Army ETL contract DACA 76-86-C-0017.

Table 1: Locations of interest points.

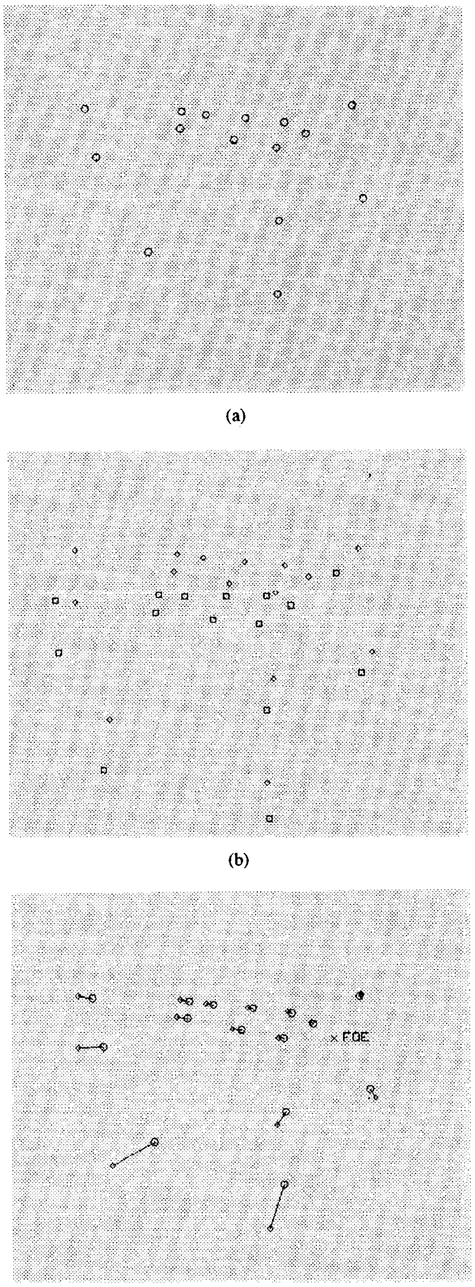
	x (ft)	y (ft)	z (ft)
1	100	25	4
2	95	-30	4
3	90	-10	4
4	85	-5	4
5	80	2	4
6	75	8	4
7	70	-8	4
8	65	10	4
9	60	0	4
10	55	5	4
11	50	-15	4
12	35	10	4
13	30	3	4
14	25	-5	4
15	20	2	4

Table 2: Location, roll, pitch, and yaw of the camera for synthetic frames A and B.

	Frame A	Frame B
x (ft)	0	5
y (ft)	0	1
z (ft)	-7	-6
roll (deg)	0	5
pitch (deg)	-15	-11
yaw (deg)	0	2

Table 3: Location, roll, pitch, and yaw of the camera for two frames of real imagery.

	Frame A	Frame B
x (ft)	x'	$x' + 0.128$
y (ft)	y'	$y' + 0.502$
z (ft)	z'	$z' - 0.0076$
roll (deg)	1.495	1.387
pitch (deg)	1.662	1.725
yaw (deg)	76.05	75.99

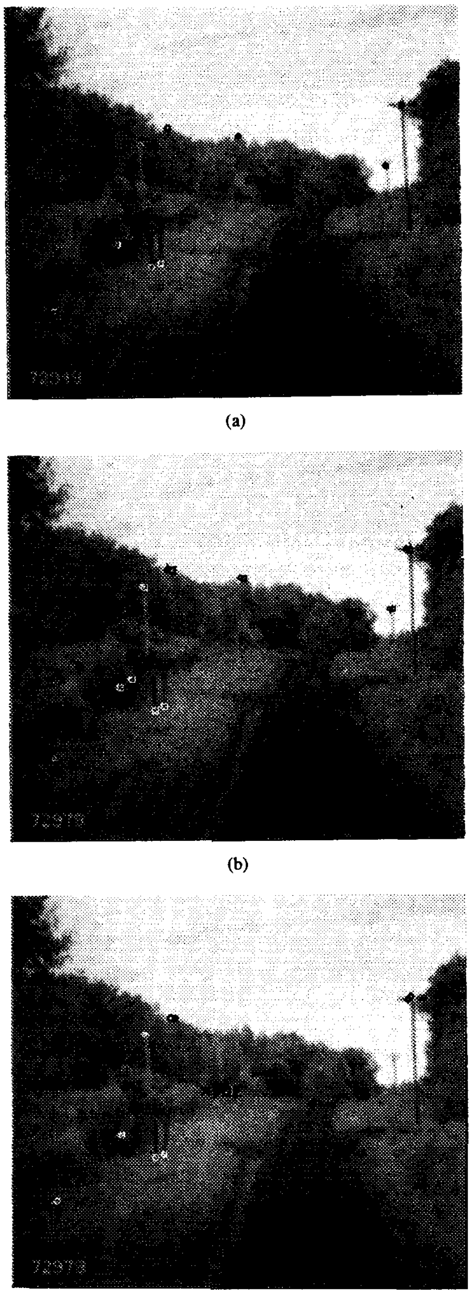


(a)

(b)

(c)

Figure 7: Optical flow results using synthetic data. *(a)* Locations of interest points in the first image, indicated by circles. *(b)* Locations of interest points in the second image, shown using squares. Diamonds indicate the derotated interest point locations. *(c)* Matching process results in displacement vectors between circles and diamonds. The FOE is indicated by a cross.



(a)

(b)

(c)

Figure 8: Optical flow results using real data. *(a)* Locations of interest points in the first image, indicated by circles. *(b)* Locations of interest points in the second image, shown using squares. Diamonds indicate the derotated interest point locations. *(c)* Matching process results in displacement vectors between circles and diamonds. The FOE is indicated by a cross.