AD A079553

INTERACTIVE AIDS FOR CARTOGRAPHY AND PHOTO INTERPRETATION

A073452

Final Technical Report
Covering the Period May 12, 1976 to October 9, 1979

SRI Project 5300
Contract DAAG29-76-C-0057

ARPA Order No. 2894-5
Program Code No. 61101E

Contract Amount: $1,373,527
Effective Date: May 12, 1976
Expiration Date: October 9, 1979

LEVEL III

By: Martin A. Fischler, Senior Computer Scientist
Principal Investigator
(415) 326-6200, Ext. 5106

and

G. J. Agin
H. G. Barrow
R. C. Bolles
L. Quam
J. M. Tenenbaum
H. C. Wolf

Prepared for:

Director
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Attention: Program Management

DDC
RECEIVED
JAN 17 1980
A

Approved for public release; distribution unlimited.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

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

DDC FILE COPY

80 1 17 033

SRI International

INTERACTIVE AIDS FOR CARTOGRAPHY AND PHOTO INTERPRETATION.

Final Technical Report. 12 May 76 - 9 Oct 79,
Covering the Period May 12, 1976 to October 9, 1979

SRI Project 5300
[redacted] DAAG29-76-C-0057,
ARPA Order [redacted] 2894-5
Program Code No. 61101E

Dec 79

13p

Contract Amount: $1,373,527
Effective Date: May 12, 1976
Expiration Date: October 9, 1979

By: Martin A. Fischler, Senior Computer Scientist
[redacted]

and

G. J. Agin,
H. G. Barrow,
R. C. Bolles
L. Quam
J. M. Tenenbaum
H. C. Wolf

Prepared for:

Director
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Attention: Program Management

Approved:

Peter E. Hart, Director
Artificial Intelligence Center

David H. Brandin, Executive Director
Computer Science and Technology Division

410 281

ABSTRACT

Research at SRI International under the ARPA Image Understanding Program was initiated to investigate ways in which diverse sources of knowledge might be brought to bear on the problem of analyzing and interpreting aerial images. The initial phase of research was exploratory and identified various means for exploiting knowledge in processing aerial photographs for such military applications as cartography, intelligence, weapon guidance, and targeting. A key concept is the use of a generalized digital map to guide the process of image analysis. The results of this earlier work were integrated in an interactive computer system called "Hawkeye." This system provides not only basic facilities necessary for a wide range of tasks in cartography and photo interpretation, but also a framework within which other applications can be readily demonstrated.

Since January 1978, work has been focused on development of a system (called the "SRI Road Expert") capable of expert performance in a specific task domain--road monitoring. This report summarizes the specific objectives, approach, and technical accomplishments of the research program.

CONTENTS

ABSTRACT ii

LIST OF ILLUSTRATIONS iv

ACKNOWLEDGMENTS v

I INTRODUCTION 1

II THE SRI ROAD EXPERT 3

A. Image-to-Data-Base Correspondence 5

B. Road Detection and Delineation 6

C. Anomaly Analysis and Classification 6

III CONCLUDING COMMENTS 9

ANNEXES

A Map-Guided Interpretation of Remotely-Sensed Imagery . . . 12

B The SRI Road Expert: Image-to-Data-Base Correspondence 13

C Detection of Roads and Linear Structures in Aerial Imagery by Computer 14

D Road Tracking and Anomaly Detection 15

E Knowledge-Based Detection and Classification of Vehicles and other Objects in Aerial Road Images 16

F Reports and Publications 17

Accession For
NTIS GRA&I
DDC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or special
A

ILLUSTRATIONS

1 SRI Road Expert 4

ACKNOWLEDGMENTS

Major contributors to the SRI Image Understanding Program were: G. J. Agin, H. G. Barrow, R. C. Bolles, M. A. Fischler, T. D. Garvey, L. H. Quam, J. M. Tenenbaum, and H. C. Wolf.

I INTRODUCTION

Research at SRI International under the ARPA Image Understanding Program was initiated to investigate ways in which diverse sources of knowledge might be brought to bear on the problem of analyzing and interpreting aerial images. The initial phase of research was exploratory and identified various means for exploiting knowledge in processing aerial photographs for such military applications as cartography, intelligence, weapon guidance, and targeting. A key concept is the use of a generalized digital map to guide the process of image analysis. Results of this earlier work were integrated in an interactive computer system called "Hawkeye."* This system, which emulates a photo interpreter's work station, provides not only basic facilities necessary for a wide range of tasks in cartography and photo interpretation, but also a framework within which other applications can be readily demonstrated.

Particular features of the Hawkeye system include a display, graphics tablet, and a natural-language interface for user communication; a map and terrain data base with facilities for answering user queries; capabilities for interactive mensuration and delineation; and the ability to automatically monitor selected sites, such as railroad yards and harbors. The paper reproduced in Annex A describes some of the capabilities developed for the Hawkeye system.

Since January 1978, work has been focused on development of a system, called the "SRI Road Expert," capable of expert performance in a specific task domain--road monitoring. Unlike Hawkeye, which was concerned with demonstrating the feasibility of a wide range of photo interpretation tasks guided by user interaction and map knowledge, the

---

* H. G. Barrow et al., "Interactive Aids for Cartography and Photo Interpretation: Progress Report, October 1977," in Proceedings: Image Understanding Workshop, pp. 111-127 (October 1977).

Road Expert is an attempt to develop a system capable of a high level of performance in a specific task. It is generally agreed that there is no existing computer system capable of human levels of performance in the analysis of imagery, and it is believed by many that such a capability would only be possible in a system capable of accessing a significant store of knowledge about world events. Our research is an attempt to determine whether a knowledge-based paradigm could be developed to achieve human-like performance in a narrowed, but still militarily and scientifically relevant, image domain.

This report summarizes the objectives, approach, and technical accomplishments of our recent research on the Road Expert. Detailed descriptions of key results are provided in annexes.

II THE SRI ROAD EXPERT

Our primary objective in developing the SRI Road Expert was to build a computer system that "understands" the nature of roads and road events. It was intended to be capable of performing such tasks as:

(1) Finding roads in aerial imagery.

(2) Distinguishing vehicles on roads from shadows, signposts, road markings, and so forth.

(3) Comparing multiple images and symbolic information pertaining to the same road segment, and deciding whether significant changes have occurred.

The system was to be capable of performing the above tasks even when the roads were partially occluded by clouds or terrain features, were viewed from arbitrary angles and distances, or passed through a variety of terrains.

To achieve the above capabilities, research was concentrated in three technical areas listed below:

(1) Image/map correspondence--Place a newly acquired image into geographic correspondence with the map data base.

(2) Road tracking--Precisely mark the center line of selected visible sections of road in the image.

(3) Anomaly analysis--Locate and analyze anomalous objects on, and adjacent to, the road surface; identify potential vehicles.

The research results were then integrated into the system depicted in Figure 1.

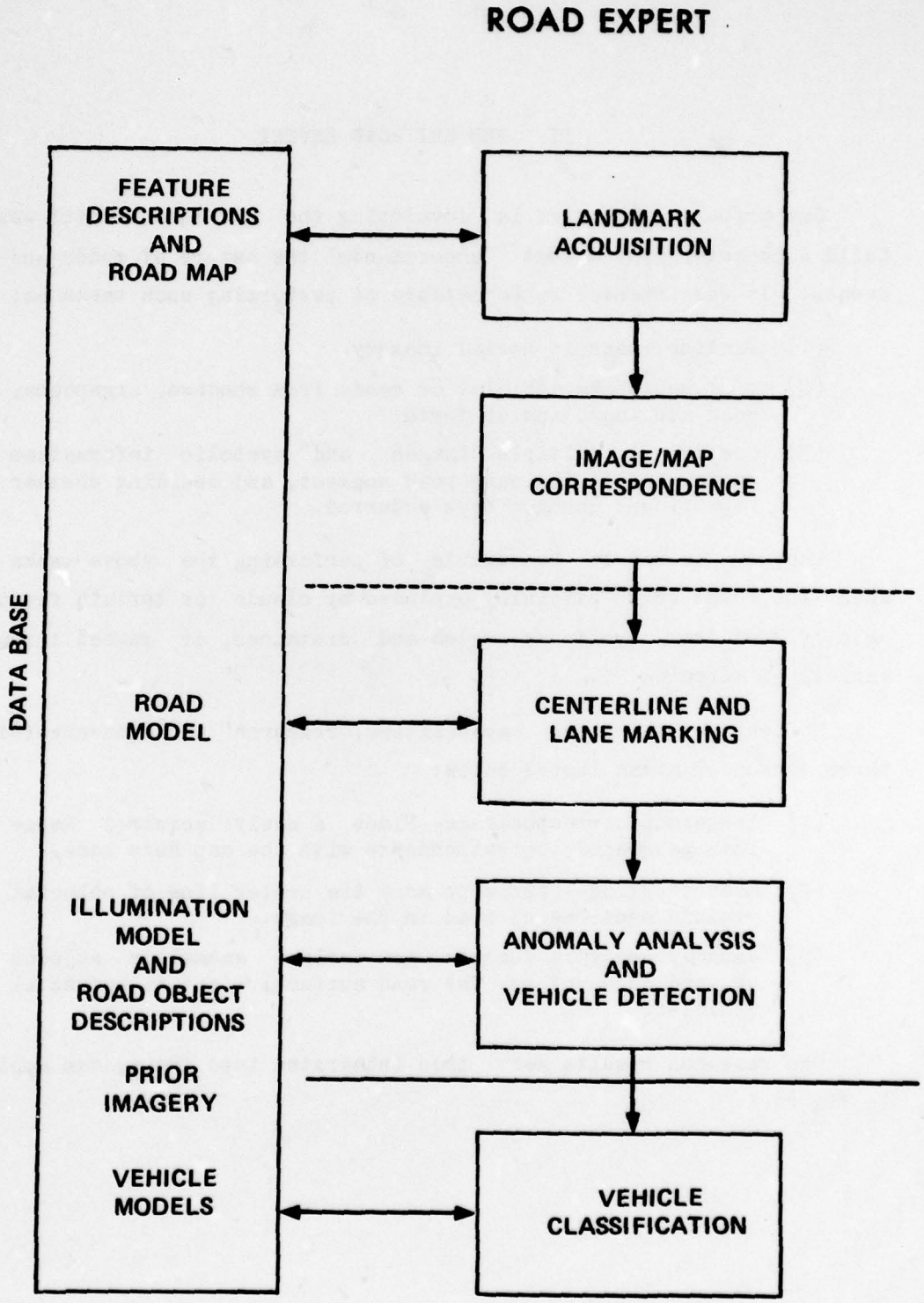


FIGURE 1 SRI ROAD EXPERT

A. Image-to-Data-Base Correspondence

The first stage in road monitoring is to establish a correspondence between the image and data base. This task involves locating a few known road features (landmarks) in a newly acquired image, and then using the correspondence between the location of these landmarks and their geographic coordinates (as stored in our map data base) to determine the precise location and orientation of the "camera" when the image was acquired. Given the camera parameters and a terrain map, we can derive a transformation that will assign geographic (x, y, z) coordinates to every point in the image. The search in the image for landmarks is a sequential process guided by our continually more precise estimate of the camera's location; as each landmark is found, we update the camera model to further reduce the search area required to locate additional landmarks.

Our work on the correspondence problem, employing an iterative approach which combines error modeling, feature matching, verification, and refinement of the camera location estimate, has resulted in a number of extensions to the existing theory.

In our experimental work it is nominally assumed that the initial combinations of uncertainties about the estimates for camera parameters imply uncertainties on the ground of approximately +/- 200 feet in X and Y for imagery with resolutions of from 1 to 20 feet/pixel. We have automatically refined the image/map correspondence to achieve an average error of between two and three feet of ground distance. Because of the accuracy and robustness of this approach, we believe that it can play an important role in an image-matching navigation or terminal homing system (e.g., the cruise missile). Details of this work are presented in Annex B.

B. Road Detection and Delineation

After the image is placed into correspondence with the map data base, one or more of the visible road sections are selected for monitoring. The road center line and lane boundaries must be located to an accuracy of one to two pixels in imagery with a resolution of one to three feet/pixel.

We developed two distinct computer-based approaches for precisely delineating roads and similar "line-like" structures appearing in aerial imagery.

At low resolution, the approach is based on a new paradigm for combining local information from multiple, and possibly incommensurate, sources, including various line and edge detection operators, map knowledge about the likely path of roads through an image, and generic knowledge about roads (e.g., connectivity, curvature, and width constraints). The final interpretation of the scene is achieved by using either a graph search or dynamic programming technique to optimize a global figure of merit. Details of this work are described in Annex C.

At high resolution, the approach is based on the use of a single coherent road model, which assumes that road segments will exhibit relatively smooth/slow changes in direction and also in intensity profile normal to road direction. A correlation-based technique is used to track the path of a road through an aerial image after it has been "acquired" (or approximately acquired) by some other means. Details of this work are described in Annex D.

C. Anomaly Analysis and Classification

The high-resolution road tracker discussed earlier assumes that roads in images are regions where brightness varies in a predictable way. Small regions in which brightness is significantly different from that predicted by the road model are called anomalies, arising from such things as vehicles, road markings, shadows of various objects on or off

the road, overhanging trees, and discolorations of the road surface. We have developed methods for detecting and classifying these anomalies. If an anomaly is judged to be a vehicle, then the program will provide a limited amount of classification as to vehicle type. If the anomaly is judged to be something other than a vehicle, the program provides the most likely interpretation of what it is.

For this task, we employed three distinct types of relevant knowledge--knowledge about the problem domain (generic knowledge), about the site (the data base), and about a particular place and time (information associated with the image).

Generic knowledge included information that could be deduced from functional descriptions (e.g., a road is a narrow, linear region upon which vehicles may travel).

Data base knowledge about the site was principally used to help locate the roads to be monitored and to allow us to predict locations of known anomalies (including both road markings and shadows of prominent off-road objects).

Image-specific knowledge, including data and time the image was acquired, allowed us to predict shadows and other photometric phenomena related to the sun's location.

In analyzing the imagery, the development of techniques for understanding shadows was crucial. Aerial scenes are often photographed in direct sunlight, and vehicles on the road cause anomalies that include the vehicle plus its shadow. Large objects off the road, such as signs, trees, and utility poles, cast shadows that are noticed by the anomaly detector. In addition, the shadows can give valuable clues as to the size and shape of the objects casting them.

We employ three basic techniques to identify shadows. A brightness model allows us to identify shadows by the absolute brightness of pixels in the difference image. A predictive model allows us to identify the portion of an anomaly most likely to be shadow when we know the position of the sun and the height of the object casting the shadow. Finally, a

projective model, which tries to locate the two long parallel sides of a vehicle, can locate the dividing line between a vehicle and its shadow.

"Expert subroutines" examine each anomaly. The vehicle expert subroutine exploits the basically rectangular shape of vehicles when viewed from above. Anomalies that are very much the wrong size are eliminated at the outset. Projecting the average brightness and average gradient magnitude upon a base line perpendicular to the presumed direction of vehicle travel enables finding the shadow and establishing a nominal width for the vehicle. Height can usually be estimated from the shadow, and length is inferred from the size of the total anomaly (allowing for a shadow fore or aft).

Two other anomaly experts, the tree shadow expert and the road marking expert, provide alternate explanations for anomalies not identified as vehicles. To qualify as a tree shadow (or the shadow of some other object off the road), an anomaly must have the appropriate average brightness, a low variance in brightness, and touch the side of the road at the side nearer the sun. Road markings (usually painted arrows or speed limit numerals) are usually brighter than the road surface, have low brightness variance, and are quite limited in extent.

Additional details of this work are described in Annex E.

III CONCLUDING COMMENTS

A central theme of this effort was to consider roads as a knowledge domain. In particular, we addressed questions of how a priori knowledge could be directly invoked by the image-analysis modules: what types of knowledge, how should it be represented, and how should it be used.

Major results produced by our overall effort are summarized below. The references in brackets refer to reports and publications listed in Annex F.

(1) The introduction and exploitation of two major paradigms:

   (a) Map-Guided image interpretation--Establishing a projective correspondence between a symbolic data base and an image, and using the data base to guide and constrain the interpretation of the image.
   [R1, R6]

   (b) Perceptual reasoning--Modeling the information sources and image operators so that selection of analysis techniques, location of search areas in the image, sequencing of information acquisition, and the combination of perceived and a priori information into a final interpretation are matched to scene content and viewing conditions.
   [R6, C9]

(2) The design and implementation of two integrated systems as frameworks for focused research:

   (a) Hawkeye--A framework for research in interactive scene analysis.
   [F1, R4]

   (b) The SRI Road Expert--A framework for understanding the requirements for achieving human-like performance in the analysis of aerial imagery. The task of road monitoring was selected as the context for this work, both for its military relevance and also to simplify the problems associated with system implementation and experimental evaluation.
   [R6]

(3) Algorithms, techniques, and representations:

(a) Image matching and image-to-data-base correspondence

* Chamfer matching and projective correspondence
[C5]

* Reprojection matching
[R3]

* Mathematical modeling of the correspondence process, and automatic determination of landmark search regions
[R6]

* Camera calibration using point-to-line landmarks
[R6]

* Verification techniques for landmark acquisition and feature detection
[R6]

* Sequential model-based strategy for establishing an image-to-data-base correspondence
[R6]

(b) Road and linear feature analysis

* Models and techniques for detecting and tracking roads in aerial imagery under a wide range of viewing conditions
[C8, R5]

* A new paradigm for combining evidence from a number of incommensurate sources (e.g., image operators), based on classifying these operators as to the types of errors they are susceptible to; and an application of this approach to the problem of linking linear segments into a composite structure
[C8]

* Techniques for detecting and classifying anomalies (including shadows, road markings, and vehicles) which appear in road scenes
[R7]

* A technique for synthetically (but realistically) introducing clouds into an aerial image
[R6]

(c) Data base structures and interactive aids to image analysis

* Mensuration techniques
  [R2, R4]

* Natural-language data base access
  [R4]

* Symbolic image data base representation for efficient geometric indexing, proximity search, and question answering (e.g., K-D Tree representation)
  [R2, R4]

Annex A

Map-Guided Interpretation of Remotely-Sensed Imagery

This paper, published in Pattern Recognition and Image Processing, describes a key contribution of the Hawkeye system--the use of map knowledge in automated photo interpretation.

MAP-GUIDED INTERPRETATION OF REMOTELY-SENSED IMAGERY

J. M. Tenenbaum, H. G. Barrow,
R. C. Bolles, M. A. Fischler, H. C. Wolf
SRI International
Menlo Park, California

ABSTRACT

A map-guided approach to interpretation of remotely sensed imagery is described, with emphasis on applications involving continuous monitoring of predetermined ground sites. Geometric correspondence between a sensed image and a symbolic reference map is established in an initial stage of processing by adjusting parameters of a sensor model so that image features predicted from the map optimally match corresponding features extracted from the sensed image. Information in the map is then used to constrain where to look in an image and what to look for. With such constraints, previously intractable remote sensing tasks can become feasible, even easy, to automate. Four illustrative examples are given, involving the monitoring of reservoirs, roads, railroad yards, and harbors.

INTRODUCTION

Aerial and satellite imagery provide an economical means of gathering large amounts of data on the earth's resources and environment. However, except in the area of survey tasks such as crop inventories and land use that can be performed with multispectral analysis, there are few economically feasible techniques for automatically extracting the useful information from such imagery.

This paper describes some initial experiments in automating an important class of remote sensing tasks that involve continuous monitoring or tracking of predefined targets. Monitoring tasks are concerned with detecting anomalous conditions at specified geographic locations. Examples include monitoring particular industrial plants for thermal or chemical pollution, oil storage facilities for spillage, forests for fires, and reservoirs for water quality. Tracking is a variant of monitoring, concerned with determining the current geographic location of a slowly moving object or boundary whose position is known approximately from a previous determination. Examples include tracking icebergs, the spreading boundaries of a known oil spill, the perimeter of reservoirs (to assess changes in water volume), coastal shorelines (to assess erosion), and the width of rivers (to assess flood threat). For such tasks, an automated system is needed that can extract updated information as new imagery arrives and distribute it directly to interested users. Multispectral analysis, by itself, is inadequate because spatial structure and context are significant factors in interpretation.

A major problem in automating such tasks is locating the designated sites in sensed imagery, that may be taken from arbitrary viewpoints. Once the image locations of a site are known, many monitoring tasks are reduced to straightforward detection or classification problems. For example, once the precise pixel location of a river passing beside a manufacturing plant is known, pollution levels in the plant's effluents can, in principle, be determined by using conventional multispectral analysis. Similarly, forest fires can be detected by looking for infrared hot spots in known forested areas. Tracking slowly changing boundaries, such as the perimeters of water bodies, is also tremendously simplified by knowledge of the boundaries' approximate prior location. Boundary detection and linking can then be accomplished using simple edge operators to verify precise edge locations along the predicted path.

To locate monitoring sites in an arbitrary image, we use a map in conjunction with an analytic camera model. The camera model is first calibrated in terms of known landmarks and then used to transform between map coordinates of designated sites and their corresponding image coordinates. By constraining where to look in an image and what to look for, a map and camera model greatly simplify the extraction of relevant information in complex aerial scenes.

MAP-IMAGE CORRESPONDENCE

A fundamental requirement in exploiting a map is to establish the geometric correspondence between image and map coordinates, which then allows known ground sites to be located in the image. Ground locations have conventionally been determined by warping the current sensed image into correspondence with a reference image, based on a large number of local correlations [1]. The reference image serves as a map indicating locations in the sensed image that correspond to previously determined points of interest in the reference image. The process is computationally expensive and limited to cases where the reference and sensed images were obtained under similar viewing conditions.

To overcome these limitations, we abandon the use of a reference image and rely instead on a symbolic reference map containing explicit ground coordinates and elevations for all monitoring sites as well as landmarks (roads, coastlines, and so forth). The geometric correspondence between this map and the sensed image is established by calibrating an analytic camera model.

CH1428-2/79/0000-0610$00.75 © 1979 IEEE

Reprinted from PATTERN RECOGNITION AND IMAGE PROCESSING 1979

A typical camera model [2] has between five and seven parameters that specify focal length and the location and orientation of the camera (in map coordinates) when the image was taken. Once these parameters are known, the image coordinates corresponding to any map location can be determined precisely with straightforward trigonometry. (The camera location and map location jointly define a ray in space. The intersection of this ray with the image plane yields the desired image coordinates.) Since image coordinates are determined for the original unrectified image, expensive image warping is unnecessary.

Map Data Base

The map data base used in this research is essentially a compact three-dimensional description of the location and shapes of major landmarks and monitoring sites. Point features, such as road intersections, small buildings, and many monitoring sites, are represented by their three-dimensional world coordinates and (where applicable) a list of characteristics to be monitored. Linear landmarks, such as roads and coastlines, are similarly represented as curve fragments with associated ordered lists of world coordinates. Ground coordinates are expressed in a standard reference frame, the UTM grid, with elevations expressed in meters above sea level. The data base can be accessed by location (e.g., What is at x, y, z?), by entity name (e.g., What is the location of factory x?), and by entity type (e.g., What factories are there?). For further details on map representation, the reader is directed to Reference [3].

Our experimental domain throughout this project was the San Francisco Bay Area, as depicted in Figure 1. Figure 2 is a computer display of a simple map data base of this area. The map contains a major landmark (the coastline) and a number of representative monitoring sites, each designated by a cross. Longitude and latitude data for the on-line map were obtained interactively from a USGS map, using a digitizing table. Elevations were read off the map and entered manually via keyboard. Although displayed as a continuous trace, the coastline, in fact, is internally represented by just 100 discrete sample coordinates.

Several map data bases, each highlighting specific features (e.g., roads, railroad yards, piers) were used in experiments described in this report. These maps have not yet been integrated into a monolithic data base, although all software necessary to do so exists (Ref. [3]).

Camera Calibration

The traditional method of calibrating a camera model requires two stages: First, a number of known landmarks are independently located in the image; and second, the camera parameters are computed from the pairs of corresponding world and image locations, by solving an over-constrained set of equations [2, 4].

The failings of the traditional method stem from the first stage: Landmarks are located in the sensed image by correlating with fragments of reference images. This requires reference images taken under the same viewing conditions as the current sensed image. Moreover, since landmarks are found individually, using only very local context (e.g., a small patch of surrounding image) and with no mutual constraints, false matches commonly occur. (The restriction to small features is mandated by the high cost of area correlation and by the fact that large image features correlate poorly over small changes in viewpoint.)

A new calibration procedure, called "Parametric Correspondence", was developed that overcomes these failings by integrating the landmark-matching and parameter solving steps and by using global shape rather than tonal appearance as the basis for matching. In this procedure, initial estimates of camera location and orientation are obtained on the basis of available navigational data. The camera model is then used to predict the appearance of landmarks in an image for this assumed viewpoint. Calibration is achieved by adjusting the camera parameters (i.e., the assumed viewpoint) until the predicted appearances of the landmarks optimally match a symbolic description extracted from the image.

A detailed description of parametric correspondence is given in Reference [5]. However, the essential ideas can be quickly grasped through an example. Figures 3-6 illustrate the process of establishing correspondence between the symbolic map of Figure 2 and the sensed image of Figure 1, using the coastline as a landmark.

First, a simple edge follower was used to trace the high contrast coastline in Figure 1, producing the edge image shown in Figure 3. Next, using initial camera parameter values (estimated manually from navigational data provided with the image), the coastline coordinates in the map were transformed into corresponding image coordinates and overlaid on the extracted edge image (Figure 4). The average mean square distance between the extracted coastline and that predicted on the basis of the assumed viewpoint was seven pixels. A straightforward hill-climbing algorithm then adjusted the camera parameters to minimize this average distance. Figure 5 shows the final state, in which the average distance has been reduced to 0.8 pixel.

Using the final parameter values, it is now possible to determine within a pixel the precise image locations corresponding to each monitoring site in the map. Only three sites are actually visible in this image: two oil depots and a coffee factory. These are shown in Figure 6, superimposed on the original image. The apparent misregistration in Figure 5 is actually the result of errors in contour extraction (Figure 3); despite such errors, the global matching criteria is still able to achieve subpixel accuracy of the projected map points. Figures 7 and 8 provide two additional examples, illustrating the ability of the calibration process to place the map in Figure 2 into correspondence with imagery taken from arbitrary viewpoints.

Parametric correspondence has some significant advantages over conventional approaches to camera

calibration that depend on reference imagery. Computational requirements (both processing and memory) are sharply reduced because a symbolic map typically contains orders of magnitude less data than a reference image. Invariance to viewing conditions (viewpoint, spectral band, sun angle etc.) is significantly improved because maps describe global shape characteristics that are relatively immune to seasonal and diurnal variation and to ambiguous matches. Moreover, since shape information is projected through the camera model before matching, distortions due to viewpoint are no longer a problem. A detailed discussion of these advantages appears in Reference [5].

MAP-GUIDED MONITORING

Having placed an image into parametric correspondence with a three-dimensional map, it is possible to predict the image coordinates of any feature in the map and, conversely, to predict the map features corresponding to any point in the image. Given this capability, many basic monitoring tasks of the type discussed in the introduction can be automated using straightforward image-analysis techniques. In Figure 8, for example, one could, in principle, test the pixels located in reservoirs for water quality, the pixels located in shipping channels beside oil depots for evidence of spillage, the pixel located at the industrial plant for evidence of particulates, and the pixel located at the Sacramento River Delta for evidence of salt water intrusion.

These examples fall within the competence of traditional multispectral analysis programs which uniformly process all pixels in an image and produce a statistical result. For such tasks, the primary advantages of map guidance are an enormous reduction in the number of pixels to be processed, potentially enhanced discrimination (resulting from the ability to optimize classification criteria at each site), and geographically specific results that are generally more useful than statistical summaries. In more complex interpretation tasks, where spatial structure and context are important, the benefits of map guidance are more profound. Four representative experiments will now be described.

Reservoir Monitoring

Consider first the problem of monitoring the water level of a reservoir. Water level, of course, is not directly measurable from an aerial image; some additional information or constraint is needed. The required information can be obtained from a terrain map in registration with the image.

As the water level rises and falls, the outline of the reservoir expands and contracts in a predictable way to follow the elevation contours of the terrain (see Figure 9). Thus water level can be determined by extracting the outline of the reservoir in the image and determining its location with respect to known elevation contours. Knowing the water level, one can then integrate over the corresponding region of flooded terrain to determine the volume of stored water. (The function relating water volume and water level is monotonic and can be tabulated for each reservoir.)

Since the surface of a reservoir is flat, the water level can be determined without a complete outline; the image coordinates of even a single point on the reservoir boundary would, in principle, suffice. In practice, elevations can be determined for a number of boundary points and averaged together to compensate for statistical uncertainties in estimating the precise image coordinates of each boundary point. (Concentrating the boundary samples where terrain slope is most gradual maximizes the sensitivity of edge location to changes in water level. See Figure 9(b).) The resulting distribution of elevations, which should be tightly clustered, provides a check on the quality of the map-image correspondence.

A reservoir monitoring procedure incorporating these ideas was implemented. First, geometric correspondence was established between the sensed image and a contour map of the terrain using the techniques described in the previous section. Correspondence was based on geographically stable landmarks unrelated to reservoir boundaries.

Second, the image coordinates of selected points on the reservoir boundary were determined to subpixel precision by analyzing the gradient of intensity along a line in the image perpendicular to the elevation contours at each point. The analysis was restricted to a contour interval bracketing the water level observed in a previously analyzed image. This constraint not only reduced computation but also served as an effective contextual filter for discriminating irrelevant intensity discontinuities, arising, for example, from other nearby bodies of water.

Third, the water level corresponding to each detected boundary point was obtained by linearly interpolating the elevations of the terrain contours used to delimit boundary detection.

Finally, the water volume corresponding to the average water level was obtained by table lookup.

Steps (2)-(4) are repeated for each reservoir in an image containing more than one.

The above procedure was tested on a set of images of Briones reservoir, the rightmost of the twin reservoirs in the upper center of Figure 8. Figure 10 is a higher resolution image of the Briones shoreline with elevation contours superimposed. The lines in Figure 11 indicate selected perpendiculars between the 500 and 550 elevation contours where the terrain slope is most gradual. The location of the land/water boundary along each of these lines was assigned to the point of maximal intensity discontinuity, as shown in Figure 12.

The water level corresponding to each boundary point was computed by interpolation. The mean water level in the present image of Briones, based on interpolating 170 boundary points, was determined to be 523.8 feet. This is within a foot of the ground-truth figure provided by the reservoir operator and corresponds to about a one percent error in volume. The accuracy of this approach is limited by the accuracy of the terrain

map, the quality of map-image correspondence, and the precision with which the land/water interface can be located in an image. These factors are discussed further in Reference [6].

Reservoir monitoring is an instance of a generic class of tasks in which it is necessary to determine the precise path through an image of a linear feature (e.g., shoreline, river, road) whose location and shape are known, perhaps only approximately, from a map. Maps can be used in such tasks to facilitate both the process of locating the boundary in the image and the subsequent interpretation of boundary characteristics in terms meaningful to an application (e.g., interpreting image coordinates as water levels). Applications of map-guided boundary verification might include monitoring river widths (and heights) for flood threat, monitoring coastlines for erosion, and monitoring river deltas for excessive silt deposit. Unlike reservoir monitoring, extensive manual ground-based monitoring is not economically feasible in these applications.

Road Monitoring

Locating known roads in an aerial image is a prerequisite for a variety of applications ranging from vehicle monitoring [7] to map updating. Finding roads is somewhat different from finding reservoir boundaries in that a thin linear feature is involved and a continuous path is needed.

Conventional sequential line-tracking algorithms are unsuitable because they are easily sidetracked whenever either the local evidence for a line is weak or other lines are present in close proximity. These contingencies arise frequently in aerial imagery because roads are usually clustered into networks and pass regularly through heavily textured areas where one or even both edges may be locally obscured.

To overcome these problems, a line-tracing algorithm was developed that uses a rough prediction of the path of a road, provided by a map, as a guide in determining the precise path. The map information constrains the analysis to relevant parts of the image and is used to bridge gaps where local pictorial evidence is weak or ambiguous. The algorithm operates by applying specially developed line and edge detectors in the vicinity of the predicted road path and then uses a parallel dynamic programming algorithm to find a globally optimal path through the local feature values. Further technical details can be found in Ref. [8].

Figures 13-16 show the tracing algorithm in action. Figure 13 is an aerial image of a rural area taken for a U. S. Geological Survey mapping project. The portion shown has been digitized into 256 x 256 pixels (representing 20-foot squares on the ground), each having one of 256 brightness levels. Overlaid on the image is a road path predicted from a map with standard (50-foot) cartographic accuracy. A local line detector was applied at all image points within a band centered on this guideline. The system then found the lowest-cost path from the start of the guideline to the finish, where the incremental path cost between adjacent image points was an inverse function of the local line detector score. The path so traced is displayed in Figure 14. Figure 15 shows the result of tracing many of the roads visible in the image. Note that the program has traced the center line of the wide road and that it has performed extremely well in areas in which the road is faint or partially obscured, such as at the lower left and the upper right of the image. Figure 16 shows the results of guided road tracing in an urban area containing many intersecting streets. The tracings have been fitted with straight line segments to cartographic accuracy. The results here, too, are extremely good.

Although we have performed only a limited number of experiments with guided tracing, the results have been most encouraging. The system is capable of tracing linear features that are hard even for a human to discern through a wide range of terrain types and environments. It needs relatively little guidance; but the more guidance it is given, the more reliable and efficient is its performance. It can accept guidance interactively (via light pen), as well as from preexisting maps. Interactive guidance is useful in map updating, allowing new roads to be carefully traced on the basis of a quick, light pen sketch.

Map-guided tracing of linear features is a requirement that arises in a variety of remote sensing tasks, for example, in the monitoring of rivers and railroad lines. Given suitable operators for detecting local evidence, the optimal path algorithm used to obtain a continuous road track should also work equally well in these other line tracing applications.

Object Verification Tasks

Railroad and highway monitoring are two examples of a generic class of remote sensing applications we shall call object verification tasks. Such tasks entail the detection, mensuration, or counting of specified entities whose possible locations and orientations in the image can be constrained by a map. The general approach is to determine the image coordinates for a reference structure (such as a railroad track, ship berth, or road) and then apply special-purpose operators to detect objects of interest (such as boxcars, ships, or cars). For example, we have implemented a boxcar-counting routine that analyzes the intensity profiles along predicted paths of railroad track in an image, looking for possible ends of trains and gaps between cars. Such events usually appear as step changes in brightness and dark, transverse lines, respectively. Hypothesized gaps and ends are interpreted in the context of knowledge about trains (e.g., standard car lengths and allowed inter-car gap widths) and about the characteristics of empty track to prune artifacts and improve the overall reliability of interpretation. The program then reports the number of cars classified by length [8]. We have also implemented a ship-monitoring program that analyzes intensity patterns alongside predicted berth locations in a harbor to distinguish ships from water. (Water characteristically has a low density of edges, [9].) Railroad monitoring is illustrated in Figure 17 and ship monitoring in Figure 18.

The key to automating both tasks lies in using a map to define a highly constrained context (i.e., area of the image) in which relatively simple tests can be used to distinguish objects of interest. Knowing the locations of tracks, for example, reduces the task of boxcar counting to a one-dimensional, template-matching problem, while knowing the locations of berths reduces ship finding to a trivial discrimination task. We believe that boxcar counting and ship monitoring are representative of a broad class of object-verification tasks that includes counting planes on runways and cars on highways, for which similar monitoring programs can be developed.

CONCLUDING COMMENTS

This paper has described a map-guided approach for automating an important class of remote sensing tasks involving long-term monitoring of predefined ground sites. The key idea is the use of a map in conjunction with an analytic camera model to constrain where to look in an image and what to look for. With map-guidance, many previously intractable monitoring tasks become feasible, in some cases even easy, to automate.

The map-guided approach has some potentially significant advantages over the exhaustive statistical style of processing currently used in applications such as crop classification. First, processing can be focused on the relevant portions of an image, sharply reducing computational costs and making feasible the use of sophisticated forms of analysis (involving texture, spatial patterns, and the like) that would be utterly impractical to apply at each pixel (16 million in a typical 4000 x 4000 LANDSAT image). Second, analysis routines can be simplified and made more reliable by exploiting knowledge of what to look for at each site. For example, classification criteria can be optimally tuned to discriminate the few relevant alternatives at each location. Finally, a map-guided analysis yields geographically specific results that are much more useful than conventional statistical summaries: Knowing that a particular factory is emitting excessive SO_2 is much more useful, for example, than knowing that 1 percent of 16 million pixels are polluted.

The practicality of automating monitoring tasks using the approach we have described depends, of course, on the availability of high resolution satellite imagery and satellite sensors that can be modeled analytically. Assuming these are forthcoming, the payoffs from automated monitoring could be substantial. We envisage systems that would extract updated information automatically as new imagery arrived and distribute it to interested users on a subscription basis. Initially, the analysis could be performed at existing ground-based data-processing facilities with only modest increases in computational load. Ultimately, the information could be extracted on-board satellites dedicated to specific monitoring functions and relayed direct to users via communication satellites. On-board processing appears feasible because of the dramatic reductions in computation made possible by the concept of map-guided image analysis.

For routine monitoring tasks with large user constituencies, centralized information extraction should significantly reduce the overheads of storing, retrieving, and distributing large volumes of data. Moreover, it would eliminate the need for installing image analysis facilities at many user sites.

REFERENCES

1. Bernstein, R., "Digital Image Processing of Earth Observation Sensor Data," IBM Journal of Research and Development, Vol. 20, No. 1 (January 1976).

2. Sobel, I., "On Calibrating Computer Controlled Cameras for Perceiving 3-D Scenes," Artificial Intelligence, Vol. 5, pp. 185-198 (1974).

3. Barrow, H. G., "Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report (Appendix A), Contract DAAG29-76-C-0057, SRI Project 5300, SRI International, Menlo Park, California (December 1977).

4. Duda, R., and P. Hart, Pattern Classification and Scene Analysis (John Wiley & Sons, Inc., New York, New York, 1973).

5. Barrow, H.G., et al., "Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching," in Proc. Fifth Intl. Joint Conference on Artificial Intelligence (Cambridge, Massachusetts, August 1977).

6. Tenenbaum, J.M., Fischler, M.A., Wolf, H.C., "A Scene Analysis Approach to Remote Sensing," SRI A.I. Center Technical Note 173, SRI International, Menlo Park, California (October, 1978).

7. Barrow, H. G., "Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, Contract DAAG29-76-C-0057, SRI Project 5300, SRI International, Menlo Park, California (June 1978).

8. Barrow, H. G., "Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, Contract DAAG29-76-C-0057, SRI Project 5300, Stanford Research Institute, Menlo Park, California (November 1976).

9. Barrow, H. G., "Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, Contract DAAG29-76-C-0057, SRI Project 5300, Stanford Research Institute, Menlo Park, California (May 1977).

ACKNOWLEDGEMENTS

This research was supported by the National Aeronautics and Space Administration under Contract No. NASW-2865 and by the Advanced Research Projects Agency under Contract No. DAAG29-76-C-0057.

Best
Available
Copy



FIGURE 1 HIGH ALTITUDE VERTICAL MAPPING PHOTOGRAPH OF SAN FRANCISCO BAY AREA

Taken from a U-2 at 45,000 feet.

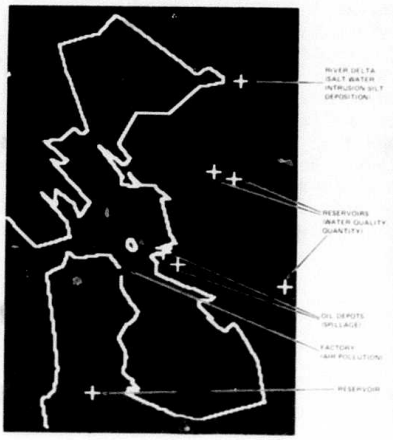


FIGURE 2 COMPUTER DISPLAY OF A SIMPLE MAP DATA BASE FOR THE SAN FRANCISCO BAY AREA SHOWING MAJOR LANDMARK (COASTLINE) AND REPRESENTATIVE MONITORING SITES (CROSSES)

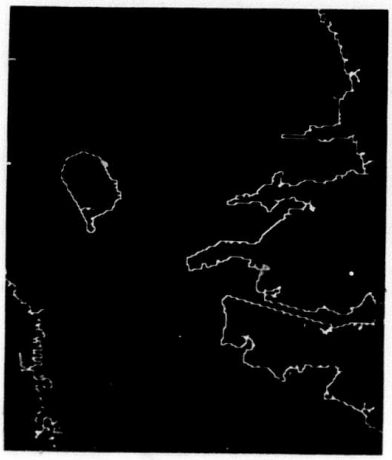


FIGURE 3 COASTLINE EXTRACTED BY BOUNDARY FOLLOWER



FIGURE 4 PREDICTED IMAGE COORDINATES OF COASTLINE, (BASED ON NAVIGATIONAL ESTIMATES OF CAMERA LOCATION AND ORIENTATION) SUPERIMPOSED ON EXTRACTED BOUNDARY



FIGURE 5 PREDICTED COASTAL COORDINATES AFTER OPTIMIZATION OF CAMERA PARAMETERS

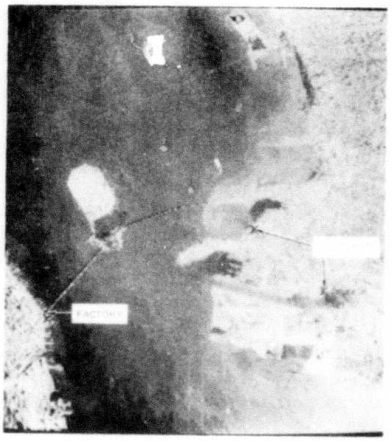


FIGURE 6 PREDICTED IMAGE LOCATIONS OF VISIBLE MONITORING SITES BASED ON OPTIMIZED PARAMETERS

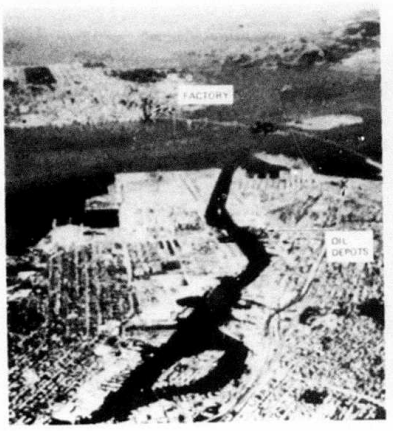


FIGURE 7 PREDICTED LOCATIONS OF VISIBLE MONITORING SITES IN AN OBLIQUE VIEW LOOKING WEST FROM ALAMEDA

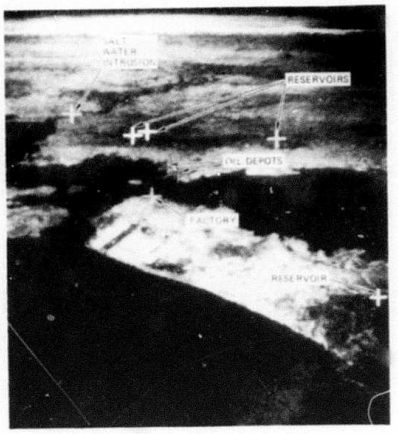


FIGURE 8 PREDICTED LOCATIONS OF VISIBLE MONITORING SITES IN A HIGH ALTITUDE OBLIQUE VIEW LOOKING EAST FROM THE PACIFIC OCEAN

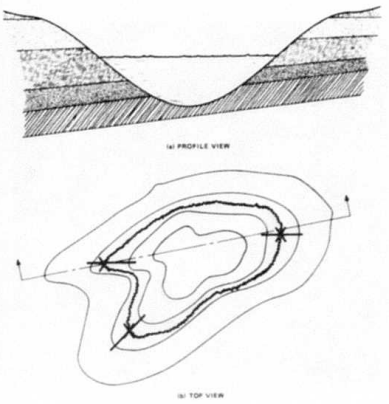


FIGURE 9 RELATIONSHIP OF WATER LEVEL TO TOPOGRAPHY OF TERRAIN

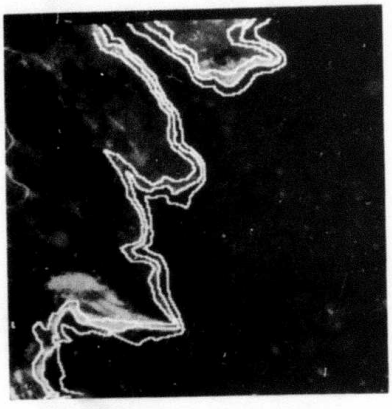


FIGURE 10 TERRAIN CONTOURS SUPERIMPOSED ON IMAGE OF BRIONES RESERVOIR.

The actual water height is 524 feet above sea level.

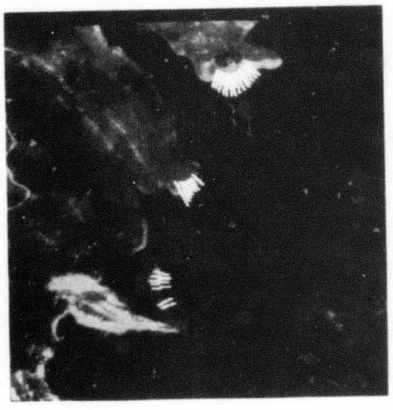


FIGURE 11 LINES DESIGNATING LOCATION FOR DETERMINATION OF LAND-WATER BOUNDARY

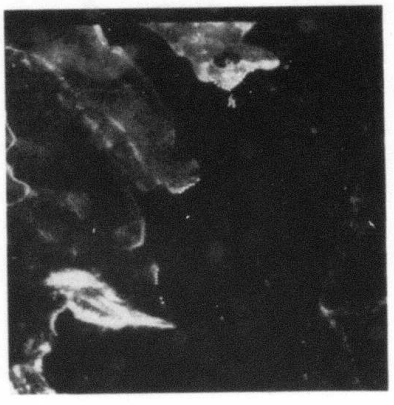


FIGURE 12 LOCATIONS OF LAND-WATER BOUNDARY ASSIGNED TO POINTS OF HIGHEST LOCAL GRADIENT ALONG LINES SHOWN IN FIGURE 11

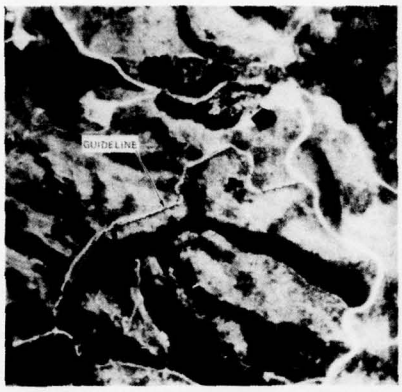


FIGURE 13 A RURAL ROAD WITH GUIDELINE

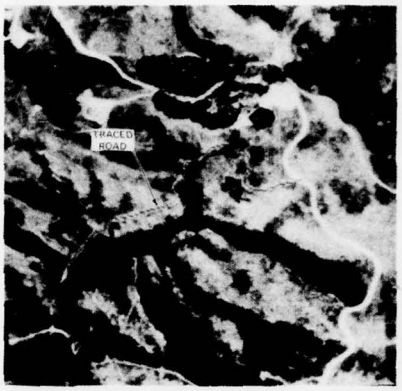


FIGURE 14 OUTPUT OF GUIDED TRACING ALGORITHM

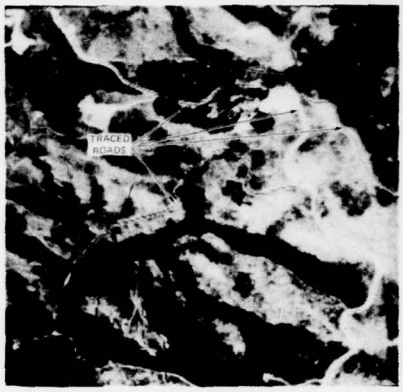


FIGURE 15 GUIDED TRACING OF SEVERAL RURAL ROADS



FIGURE 16 GUIDED TRACING OF SEVERAL URBAN STREETS



FIGURE 17 AUTOMATED BOXCAR COUNTING

Lines indicating track locations were traced interactively in this example but could have been obtained by putting image in correspondence with a three-dimensional map of the railyard, as in the ship example of Figure 18. Statistical operators are flown along tracks to detect dark transverse lines that are characteristic of gaps between boxcars. Boxcars are indicated by dots whenever the spacing between hypothesized gaps is consistent with knowledge of standard car lengths.

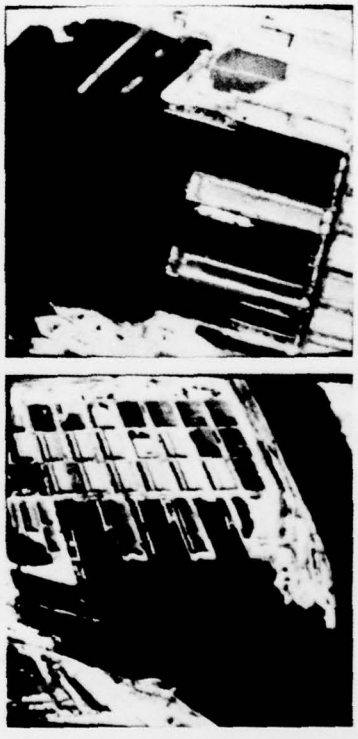


FIGURE 18 AUTOMATIC SHIP MONITORING

The guidelines indicating known berth locations were obtained for both images from the same three-dimensional map of Oakland Harbor, based on determination of viewpoint for each image. The light, wiggly lines beside the berths indicate regions of high edge content, characteristic of ships.

Annex B

The SRI Road Expert: Image-to-Data-Base Correspondence

(From "Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, SRI Project 5300, October 1978, pp. 11-52)

II THE SRI ROAD EXPERT: IMAGE-TO-DATA BASE CORRESPONDENCE

A. Introduction

Computing an image-to-data base correspondence is a general problem occurring in all knowledge-based systems. In most image tasks the correspondence is a projective transformation and can be modeled as a function of the camera parameters, such as focal length, X, Y, Z, heading, pitch, and roll. If the parameters are known precisely, the model can precisely predict the two-dimensional image coordinates for any three-dimensional data base point.

One common form of the image-to-data base correspondence problem is to be given good estimates of the camera parameters and be asked to improve them. This task is important in many military situations. For example, in navigation it is the crucial step that improves the system's estimate of the location of the plane or missile. In change detection it is used to align two images of the same area so that the corresponding regions can be compared. In the Road Expert it is the key to the utilization of the data base in subsequent tasks such as road monitoring.

The basic approach we are using to refine a correspondence is to locate known features in the image and use their locations to improve the correspondence (see Figure 9). The data base contains descriptions of the available features. From these descriptions a set of features is chosen to be located that is based on the predicted viewpoint and viewing conditions. The estimates of the camera parameters are used to predict what the features look like and where they are likely to appear. Feature detection techniques ("operators") are chosen to locate the features and they are applied. Since the operators may not locate their intended features, their results are verified either by locating a larger portion of the features or by checking the relative positions of

other features. After a set of features has been found, their locations are used to refine the estimates of the camera parameters. The parameters are refined by searching the parameter space for sets of parameter values that minimize the distances between the predicted locations of features and the locations determined by the operators. If the correspondence is not precise enough, the whole process can be repeated.

The important computations and decisions required to refine a correspondence are listed below:

(1) selection of features
(2) prediction of the appearance of a feature
(3) selection of an operator to locate the feature
(4) prediction of the nominal image location of a feature
(5) prediction of the range of image locations about a feature's nominal location
(6) selection of the order in which to apply the operators
(7) application of the operators
(8) verification of the results produced by an operator
(9) decision of when to use the results of one or more operators to help other operators locate their features
(10) decision of when to update the whole correspondence
(11) computation of a refined correspondence
(12) decision to stop

A number of people have worked on individual items in this list [1, 5, 6, 7, 8, 9, 10, 11, and 12], but mainly for pairs of images that were taken closely in time and from similar viewpoints.

There are several factors in the military domain, as well as other domains, that increase the difficulty of these items beyond current capabilities. Examples of such factors are a wide variety of viewpoints, a distribution of shadows, and the possibility of clouds. All of them make it more difficult to select features, predict the appearance of features, and locate features. Therefore, they increase the need for feature verification and strategy-based decisions. Which

operators should be used for an image taken from this viewpoint and under these conditions? When should the results of one operator be used to reduce the predicted search area for a nearby feature? This type of question becomes more important as features become harder to find.

Our research goal is to produce an automatic system to refine correspondences within the road domain. To reach this goal we need to develop new models and techniques for several of the items in the above list. So far we have concentrated on a few of them: the prediction of the range of image locations for a feature, the verification of the results of an operator, and the computation of a refined correspondence. In this section we will state our assumptions, describe our new techniques, and present an example.

B. Assumptions

Our assumptions are summarized in Figure 10.

Figure 11 is a typical picture to be processed by the system. We assume that the resolution of the digital images will be between 20 feet/pixel and 1 foot/pixel. Figure 12, which is another picture of the site shown in Figure 11, is displayed so that one pixel corresponds to approximately sixteen feet on the ground. Figure 13 is a portion of Figure 11 displayed at its full resolution of approximately 1 foot/pixel.

We assume that we will have a data base of the area on the ground contained in each picture to be analyzed. The data base contains the geometry and topology of the roads and the locations of other features, such as road markings. Since we expect to obtain repetitive coverage of the areas of interest, the data base may also contain information about the appearances of the road sections and features derived from previous images.

Images of the same site may be taken at different times of the day so the shadows may be different. Notice the variation in shadows between Figures 11 and 12. Part of the information expected by the

system for each picture is the day of the year and the time of day at which the picture was taken.

Some of the images may contain clouds that obscure some of the roads and other data base features (e.g., see Figure 14); and more generally, terrain features, buildings, and trees may obscure features of interest. The implication is that the system should be able to handle operators that find multiple matches, incorrect matches, or no matches at all.

Different pictures of the same region may be from different viewpoints. In particular, they may be from significantly different altitudes (e.g., twice as high) or different angles (e.g., 45-degree obliques versus vertical pictures). Figures 11 and 12 are pictures of the same site except that Figure 12 was taken from approximately twice the height and at a heading that is different from that of Figure 11 by almost 90 degrees. The wide variety of viewpoints implies that intensity correlation is not always sufficient to locate features. Other operators will be necessary.

Even though the viewpoint may vary widely, we expect to be given good estimates of the camera parameters for each picture. The camera parameters can be factored into two convenient sets: internal camera parameters and external camera parameters. The internal parameters describe the camera-specific information, such as the focal length of the lens. The external parameters describe the relative position and orientation of the camera with respect to the world represented in the data base. Generally, the a priori estimates of the internal parameters are much better than the estimates of the external parameters.

We expect a measure of the uncertainty associated with each parameter estimate. For example, the HEADING might be estimated to be 75 degrees, plus or minus one degree. These uncertainties are used to predict the regions in a picture to be searched in order to locate a feature. We will refer to these search regions as "uncertainty regions." The smaller the uncertainties, the smaller the uncertainty regions; the smaller the uncertainty regions, the easier it is to automatically locate the desired features.

Two of our most important assumptions restrict the range of initial uncertainties about the camera parameter estimates. The first one restricts the combined internal and external uncertainties so that they do not imply uncertainty regions on the ground of more than approximately plus or minus 200 feet. The second one restricts the size of each parameter's uncertainty so that it is relatively small. The first assumption, in effect, restricts the sizes of the uncertainty regions that have to be searched to locate a feature. For example, if an image has a resolution of 1 foot/pixel, the largest uncertainty region would then be approximately 400 x 400 pixels. The second assumption limits the portion of the parameter space that the optimizer has to search. It also indirectly limits the maximum geometric change in the appearance of a feature.

An implicit assumption behind the characterization of a correspondence as a function of the camera parameters is that the imaging process can be modeled as a perspective transformation. If it cannot, a different mapping function would have to be used, but the same numerical approach would apply.

C. Uncertainty Regions

Given parameter estimates and uncertainties about those estimates, where in the image is a feature likely to appear? Or more specifically, what region in the picture will have a given probability (e.g., a 95% probability) of containing the feature? To answer this question, one has to predict the effect on the location in the image of a feature caused by changing the parameter values in accordance with their stated uncertainties. To do that, one needs a model of their uncertainties. The error model we use is that the parameters vary according to a joint normal distribution, which is a reasonable assumption for measurements produced by a device such as an inertial guidance system because each parameter's error is a sum of several small errors. For this model the uncertainty regions are ellipses in the image plane. The derivation of this fact can be found in Appendix A.

Figure 15 shows a typical uncertainty ellipse that is prescribed to have a 95% probability of containing the actual occurrence of the feature. The 100 dots were produced by varying the camera parameters 100 different times according to the error model and by projecting the three-dimensional feature point onto the image plane containing the ellipse. Notice that 92 of the points are inside the ellipse, which is consistent with the 95% prediction.

Having found one feature, one would expect that its location would greatly restrict the possible locations for a nearby feature. This idea leads to a second type of uncertainty region, a relative uncertainty region. In addition to the normal information used to compute an uncertainty region, a relative uncertainty region is a function of another feature and its location. Since the location of a nearby feature typically adds constraints on the possible locations for a feature, the relative uncertainty region is usually significantly smaller than the regular uncertainty region. Given the assumption that the camera parameters vary according to a joint normal distribution, the relative uncertainty regions are also ellipses. A derivation of the mathematical description of a relative uncertainty region is given in Appendix B.

A relative uncertainty region is used to reduce the amount of work required to locate a second feature after a nearby feature has been found. This is particularly useful when a possible match for a feature is being verified. The logic is as follows: if this is feature A, then feature B should be in a small region over there; if B is not there (and not occluded), this must not be A.

Figure 16 shows the initial uncertainty ellipse and the relative uncertainty ellipse about a point feature. The large ellipse is the uncertainty region predicted from the uncertainties about the camera parameters. The small ellipse is the relative uncertainty region derived from the location of the arrow just above it in the picture.

D. Point-on-a-Line Matches

Almost all previous work has involved the use of point-to-point matches to refine correspondences. Since roads are the major objects of interest for the road expert, we wanted to include them as features that could be used within the image-to-data base correspondence phase as well as in the monitoring phase.

There is a built-in trade-off between point features and line features, such as roads: it is easier to find a point on a line than it is to locate a point feature, but less information is gained by doing so. Point-to-point matches produce twice the number of constraints for the refinement process, but they are generally more expensive to find because an area search is required as opposed to a linear search for point-on-a-line matches.

To use linear features we needed an operator (or operators) to find points on roads and we had to to extend the correspondence refinement process to include the new type of feature match.

1. Point-on-a-Line Operators

Currently we have two operators that locate points on a road. One is used at low resolution (e.g., 20 foot/pixel) when roads appear as lines, and one is used at high resolution (e.g., 1 foot/pixel) when the internal structure of the road is discernable. The low-resolution operator is an extension of the Duda road operator, which has been discussed in previous SRI image-understanding reports [2]. The high-resolution operator is an adaptation of Quam's road tracking operator [12]. It performs a 1-D correlation of the expected road cross section to locate possible points on the road and then tries to track the road for a short distance to make sure that the candidate point is part of the expected road.

2. Correspondence Refinement

The correspondence refinement process (or "optimizer") is based on Gennery's approach to calibration [11] (see Appendix C). It solves the nonlinear problem by iteratively solving linear approximations. For point-to-point matches a 3-D point in the world is matched with a 2-D point in the image. In that case the optimizer has two residuals per match to use to improve the camera parameter estimates: the X and Y components of the difference between the predicted image of the world point and the point in the image at which the operator located its match. If instead of locating a specific point, an operator locates a point on a line, the optimizer only has one residual to use because the point could be any place along the line. The residual for a point-on-a-line match is the distance from the point to the line. As the optimizer searches for improved camera parameters, the image of the 3-D line should get closer to the point located by the operator, but the closest point on the line may slip back and forth along the line.

So far the optimizer has only been extended to handle point-on-a-line matches. However, since roads are generally constructed as combinations of linear segments and arcs of circles, it may be useful to extend the optimizer to include other types of matches that involve a point and an analytic curve, e.g., a point-on-an-ellipse match. The main components of such an extension are (1) a procedure to compute the distance between a point and the curve and (2) a procedure to compute the partial derivatives of that distance with respect to the camera parameters.

The optimizer could even be extended to arbitrary curves by incorporating a procedure, such as chamfering [5], that computes the distance between a point and an arbitrary curve.

The current implementation of the optimizer is relatively fast. It takes one second on our KL-10 to perform one iteration when 100 residuals are used to refine the estimates. (Recall that each point-to-point match adds two residuals; each point-on-a-line match adds

one residual.) Five to ten iterations are normally required to achieve convergence, which is defined to be a state in which the parameter adjustments are on the order of .00005 units.

As Gennery points out, the optimizer can be used to filter out "mistakes" by iteratively deleting the match with the largest residual until the deletion no longer significantly improves that point's residual. In practice this heuristic has proven to be useful, but it is expensive and theoretically unsound. For example, consider Figure 17, which shows a set of points through which a line is to be fitted using a least-squares approach. The one "mistake" happens to draw the line toward it in such a way that the point with the worst residual after convergence is one of the "good" points. Deleting the point with the worst residual and trying again only repeats the situation. The conclusion is to try to filter out mistakes before they are given to the optimizer. The next subsection describes some of the ways this filtering or verification can be done.

E. Feature Verification

As mentioned in the last subsection, it appears to be more cost-effective to filter out mistakes, if at all possible, before applying the optimizer. We have identified four possible methods for performing such filtering:

(1) Operator threshold--Be suspicious of any match for which the operator does not produce a confidence above a certain threshold; e.g., if a 2-D correlation operator produces a correlation of less than .8, ignore its results.

(2) Self-support--Be suspicious of any match that cannot be verified by locating a larger portion of the same feature; e.g., if an operator locates a point that is supposed to be on a road but the road tracker cannot extend the match, ignore it.

(3) Pairwise support--Be suspicious of any match that is not positioned correctly relative to some other feature that has already been located; e.g., if an operator locates an arrow on a road and its matching location is not at a reasonable distance from another nearby feature that has been verified, ignore the match.

(4) Group support--Be suspicious of any match that is not positioned correctly relative to a group of other features that have already been located, e.g., if three point features have been found and verified, ignore a match for a fourth feature that does not appear at the correct relative location.

We differentiate between these methods (or heuristics) because they generally require different models and techniques.

It is relatively straightforward to apply all of the verification methods to point features. The relative uncertainty regions can be used to determine if two features are mutually consistent. This pairwise consistency can be extended to group consistency through maximal clique techniques [1] or through optimal embedding techniques [9].

The extension to group consistency can be achieved by constructing a graph that has one node for each match and a link between each pair of nodes that is pairwise consistent. The largest completely connected subgraph (i.e., the largest maximal clique) represents the largest set of mutually consistent matches. Any match that is not in that set is pairwise inconsistent with at least one of the matches in the set. Thus, it is suspicious.

Additional care has to be taken to apply the verification techniques to point-on-a-line matches. The important test is to be able to distinguish pairwise consistent matches from pairwise inconsistent matches when one or more of the matches is a point-on-a-line match. Figure 18 shows the three significantly different cases. In Figure 18a one of the two matches is a point-to-point match and one is a point-on-a-line match. If the slope of the line is known accurately, the distance between the point and the line can be used to determine if the matches are consistent. Since the uncertainties associated with each camera parameter are relatively small, the slope of the line should remain relatively constant. Thus the distance from the point to the line should be relatively constant.

In Figure 18b both of the matches are point-on-a-line matches, and the lines are essentially parallel. In this case the distance between

the lines is sufficient to check the relative positions of the two matches. For example, if an operator is trying to locate both sets of lanes on a freeway, the distance between the two sets of lanes should be within a predetermined range.

If both of the matches are point-on-a-line matches and the lines are not parallel, as in Figure 18c, some additional information is needed in order to check their relative consistency. One solution is to intersect the two lines and use that point in conjunction with a third match to check the relative position of all three matches.

F. Example

We have implemented one fixed strategy in terms of the verification techniques and are just beginning to explore the possibility of automatically tailoring the verification strategies to fit specific sets of features and tasks. The example task is to refine the image-to-data base correspondence for the picture shown in Figure 12 using its full resolution of approximately 2 feet/pixel. The initial uncertainties about the camera parameters imply uncertainties in the image of plus or minus 95 pixels, which correspond to approximately plus or minus 190 feet on the ground. The goal is to reduce these uncertainties to approximately plus or minus one pixel, an increase in precision of almost two orders of magnitude.

The data base used in this example contains two types of features, linear road segments and road surface markings. Figure 19 shows the locations of features that are available for this site. The lines represent the road segments and the pluses represent the surface markings. The appearance of each road segment is described by a road cross section model. The appearance of a surface marking is described by an image patch from a previous picture of the site.

A fixed strategy has been implemented to use these features to perform the task and demonstrate our new techniques. The basic approach is to locate the linear features first because they are less expensive to find, use them to refine the camera parameters, locate the point

features, use them to verify the first refinement, and then perform a second refinement using both the points and the lines.

Given estimates for the camera parameters, the system predicts the location of the road segments in the new picture. Figure 20 shows these predictions, which are shifted left and down approximately 60 pixels from their actual locations. The estimates of the camera parameters are also used to warp each road cross section to the expected size and orientation of the corresponding road segment. In addition, the estimates of the uncertainties about the camera parameters are used to predict the uncertainty regions about the center points of each linear segment. Figure 21 shows those uncertainty ellipses that have a 95% probability of containing the desired point.

The search strategy for a linear feature is to look along lines perpendicular to the expected location of the feature. The lengths of the lines are determined by the size of the uncertainty ellipse.

The high-resolution, one-dimensional correlation operator is applied along the search line to locate points that may be on the desired road. The self-support method is used to verify each candidate point. The road tracker tries to track the road for a short distance. If it cannot, the point is abandoned. Figure 22 shows an example of the application of self-support. The line on the left is the predicted location of the road segment. The other line, which is crossed like a T, represents the location of the match and the results of the road tracker following the road.

For some road segments self-support is not sufficient to locate the desired road because there are two or three parallel roads that all look alike. In order to distinguish one road from another, preplanned groups of features have been established within which pairwise and group support can be obtained. For example, Figure 23 shows a set of three sets of lanes, two of which are difficult to tell apart simply by looking at their road cross sections. The relative locations of the three sets of lanes are used to determine the correct matches. The lines perpendicular to the roads indicate the final choice for a consistent set of matches.

Figure 24 shows the results of searching for all of the road segments in the data base (shown in Figure 19). Two of the roads were not found because the contrasts were not sufficient to produce matches with the desired confidence. The matches were given to the optimizer along with the initial estimates of the camera parameters and the uncertainties about the estimates; the optimizer produced new estimates for the parameters and new uncertainties. Figure 25 shows the new predictions for the locations of the road segments. The new uncertainties imply uncertainties in the image of approximately plus or minus 1.5 pixels, close to our goal.

To verify the new estimates the surface markings were located. The new estimates were used to predict the locations and appearances of the features; the new uncertainties were used to predict the uncertainty regions; and two-dimensional correlation was used to locate the features. The average difference between the predicted location and the matching location was approximately 1.3 pixels, and the largest distance was 1.7 pixels. The final refinement based on both the lines and the points reduced the uncertainties in the image to approximately 1.1 pixels, which is very close to our goal and corresponds to approximately 2.2 feet on the ground.

We have begun to experiment with pictures containing clouds that obscure some of the features to be used for calibration. For example, consider Figure 26 in which several of the road segments are partially occluded. Figure 27 shows the linear features that the system could find and verify.

G. Discussion

We have described and demonstrated a set of techniques to perform some of the subtasks required in an automatic system to refine image-to-data base correspondences. In particular, we discussed techniques to compute uncertainty regions, techniques to incorporate point-on-a-line matches, and techniques to verify the results of operators. These techniques were combined to form a strategy, which we demonstrated in an example task.

Additional research is required on several other key subtasks required in an automatic system; for example, the selection of features and the tailoring of a strategy to different tasks. Other needs include better feature modeling, better operators to locate features over a wide range of viewing angles and conditions, and an alternative to least-squares optimization.

REFERENCES

1. A.P. Ambler et al., "A Versatile Computer-Controlled Assembly System," Proc. Third IJCAI, pp. 298-307 (August 1973).

2. H.G. Barrow, "Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, SRI Project 5300, SRI International, Menlo Park, California (November 1976).

3. H.G. Barrow, "Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, SRI Project 5300, SRI International, Menlo Park, California (October 1977).

4. H.G. Barrow et al., "Interactive Aids for Cartography and Photo Interpretation: Progress Report, October 1977", PROCEEDINGS: IMAGE UNDERSTANDING WORKSHOP", October 1977, pp. 111-127.

5. H.G. Barrow et al., "Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching," Proc. Fifth IJCAI, pp. 659-663 (August 1977).

6. R.C. Bolles, "Verification Vision for Programmable Assembly," Proc. Fifth IJCAI, pp. 569-575 (August 1977).

7. B.L. Bullock et al., "Finding Structure in Outdoor Scenes," Hughes Research Report 498, (July 1976).

8. L. Davis, "Shape Matching Using Relaxation Techniques," TR-480, Computer Science Dept., University of Maryland, College Park, Maryland (September 1976).

9. M. Fischler and R. Elschlager, "The Representation and Matching of Pictorial Structures," IEEE Trans. Comp., No. 22, pp. 67-92 (1973).

10. T.D. Garvey, "Perceptual Strategies for Purposive Vision," Technical Note 117, SRI International, Menlo Park, California (September 1976).

11. D.B. Gennery, "A Stereo Vision System for an Autonomous Vehicle," Proc. Fifth IJCAI pp. 576-582 (August 1977).

12. B.K.P. Horn and B.L. Bachman, "Using Synthetic Images to Register Real Images with Surface Models," Proc. Image Understanding Workshop, pp. 75-95 (October 1977).

13. L.H. Quam, "Road Tracking and Anomaly Detection in Aerial Imagery," Proc. Image Understanding Workshop, pp. 51-55 (May 1978).



FIGURE 1 OVERVIEW OF THE PM280 SITE

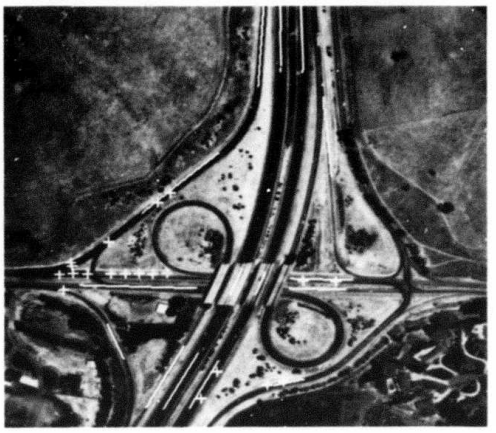


FIGURE 2(a) LOCATION OF PM280 SITE LANDMARKS

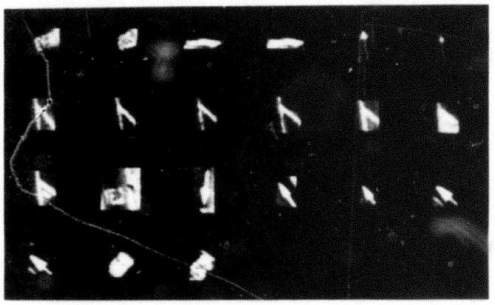


FIGURE 2(b) ROAD SURFACE MARKINGS USED AS "POINT" LANDMARKS

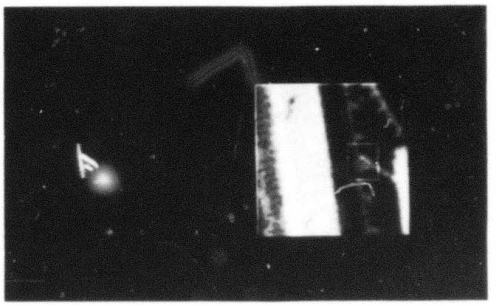


FIGURE 2(c) A POINT LANDMARK AND ITS APPEARANCE IN AN IMAGE



FIGURE 3 UNCERTAINTY ELLIPSES FOR LOCATING A KNOWN LANDMARK

The Larger Ellipse Represents the Initial Uncertainty in Locating a Road Surface Landmark. The Small Ellipse is the Refined Estimate of Location after One Other Nearby Landmark Has Been Located.

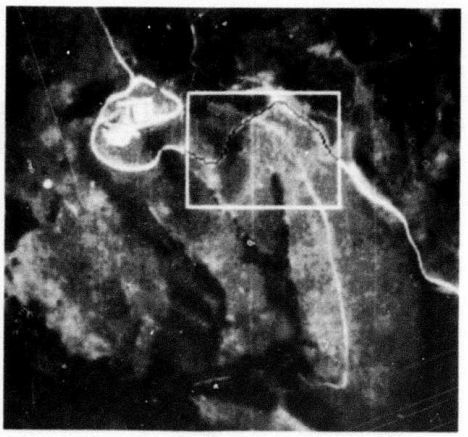


FIGURE 4 A ROAD LOCATED AND MARKED IN A SPECIFIED SEARCH WINDOW BY THE LOW RESOLUTION ROAD TRACKER

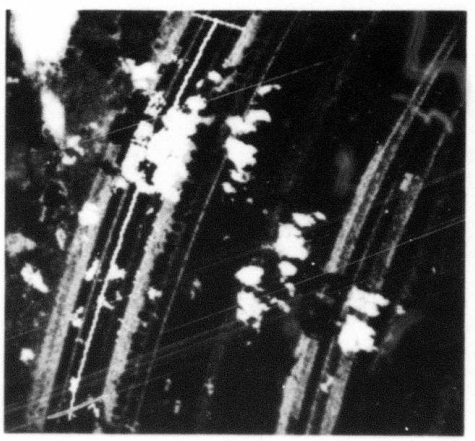


FIGURE 5(a) THE HIGH RESOLUTION ROAD TRACKER FOLLOWING A ROAD IN THE PRESENCE OF CLOUD COVER



FIGURE 5(b) THE HIGH RESOLUTION ROAD TRACKER FOLLOWING A DIRT ROAD

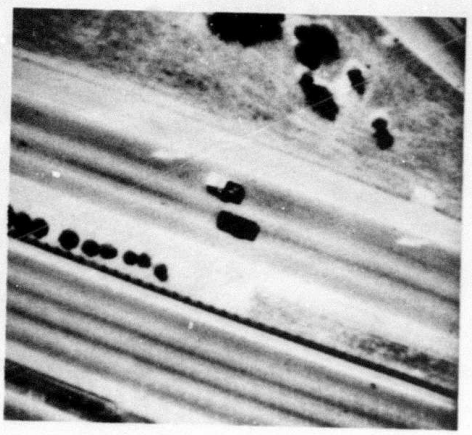


FIGURE 6(a) ORIGINAL SEGMENT OF AN IMAGE



FIGURE 6(b) DETECTION OF ANOMALOUS AREAS ON THE ROAD SURFACE

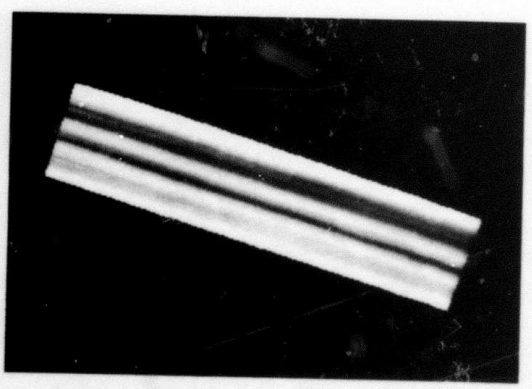


FIGURE 6(c) INTENSITY MODEL OF THE ROAD SURFACE

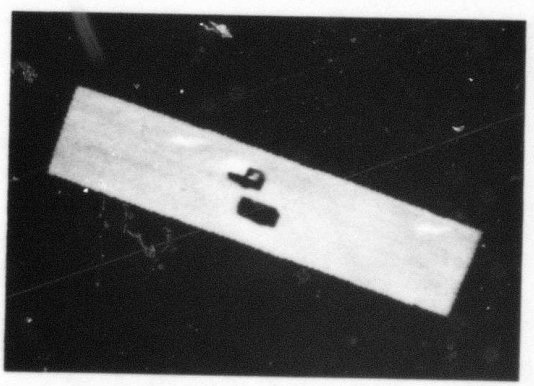
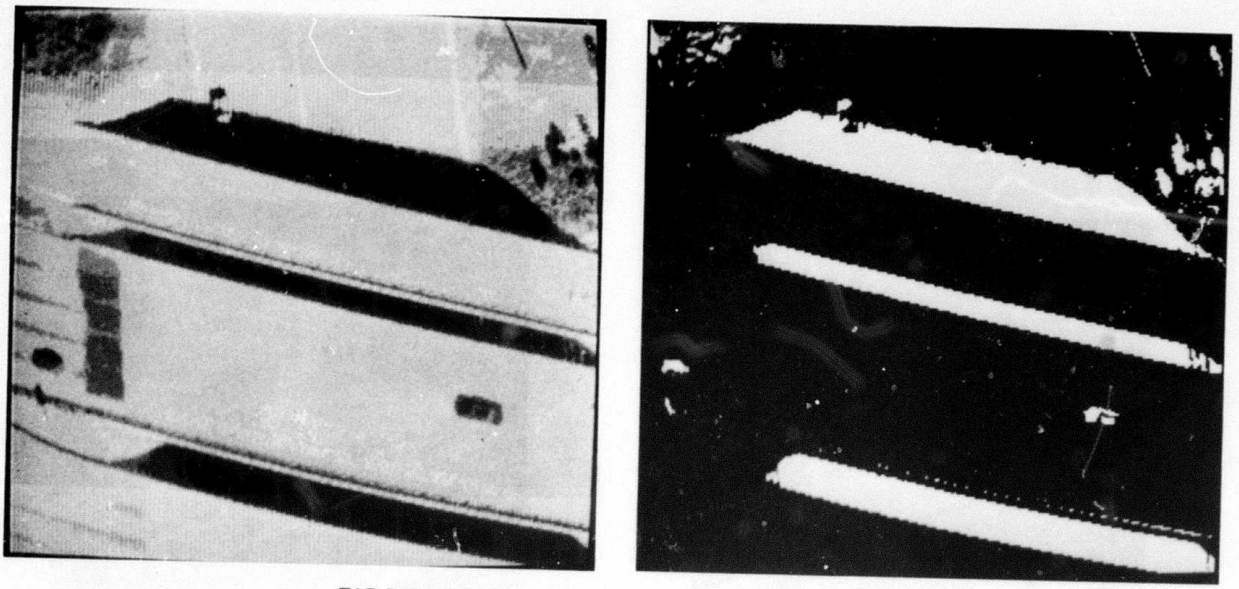
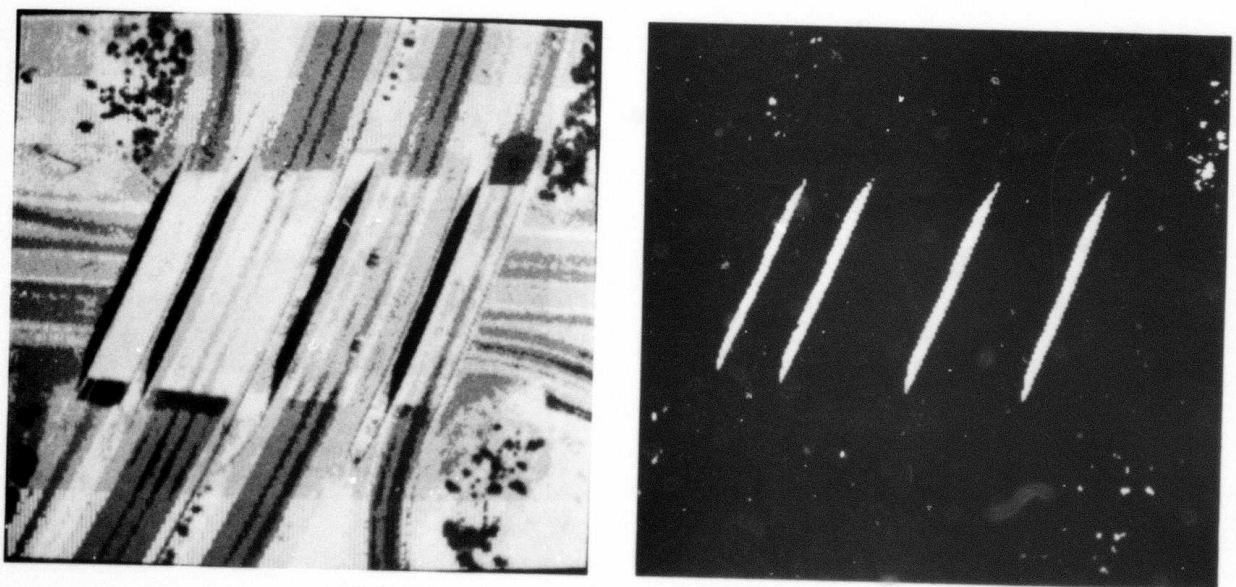


FIGURE 6(d) SUBTRACTION OF NOMINAL ROAD SURFACE INTENSITIES TO ENHANCE ANOMALIES FOR FURTHER ANALYSIS



TAR PATCH = DARK VERTICAL PATCH AT LEFT CENTER



TAR PATCH = DARK PATCH ABOVE RIGHTMOST OVERPASS

FIGURE 7(a) SHADOW EXTRACTION — THRESHOLD SET TO VALUE BELOW INTENSITY OF TAR PATCH

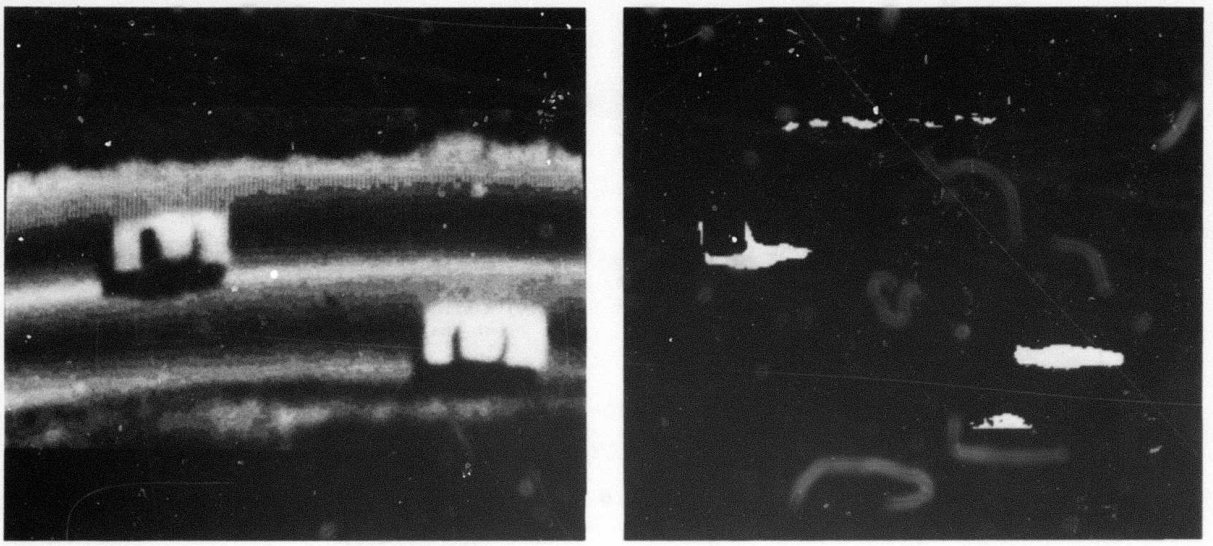
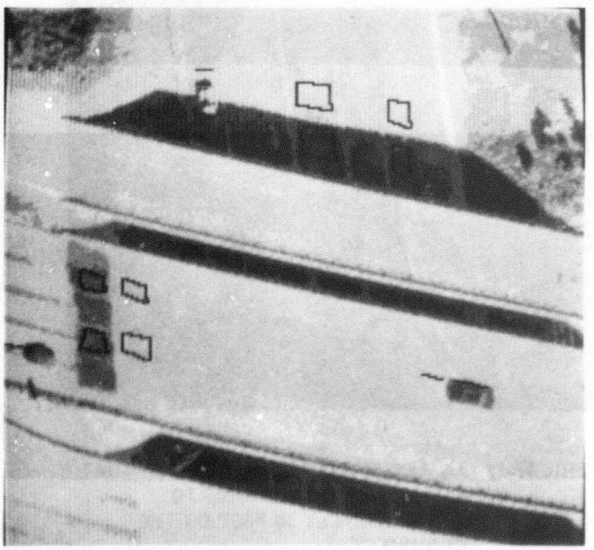


FIGURE 7(b) SHADOW EXTRACTION — THRESHOLD SET TO VALUE BELOW INTENSITY OF OIL SLICK IN THE MIDDLE OF UPPER LANE



OVERPASS SHADOWS (2 PATCHES)	0.74, 0.73
CAR SHADOWS (3 CARS)	0.79, 0.79, 0.80
TAR PATCH (2 PATCHES)	0.85, 0.86

FIGURE 8 SHADOW BOUNDARY INTENSITY RATIOS

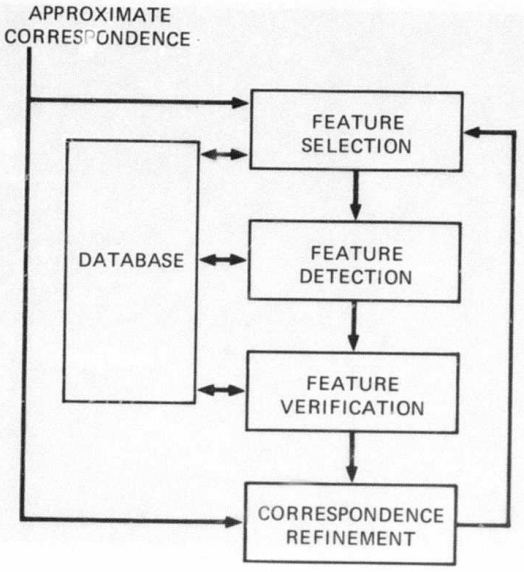


FIGURE 9 THE BASIC CORRESPONDENCE REFINEMENT PROCESS

GENERAL ASSUMPTIONS

(1) Road pictures
(2) Repetitive coverage
(3) Ground resolutions between 20 feet/pixel and 1 foot/pixel
(4) Database of roads and other features
(5) Different sun angles
(6) Database features may be obscured by clouds, terrain features, etc.
(7) Wide range of viewpoints
(8) Correspondence is a perspective transformation
(9) Small parameter uncertainties
(10) Maximum uncertainty regions on the ground of +-200 feet

INFORMATION FOR EACH IMAGE

(1) Internal camera parameters (estimates & uncertainties)
(2) External camera parameters (estimates & uncertainties)
(3) time of day and day of year image was taken

FIGURE 10 THE CORRESPONDENCE TASK ASSUMPTIONS

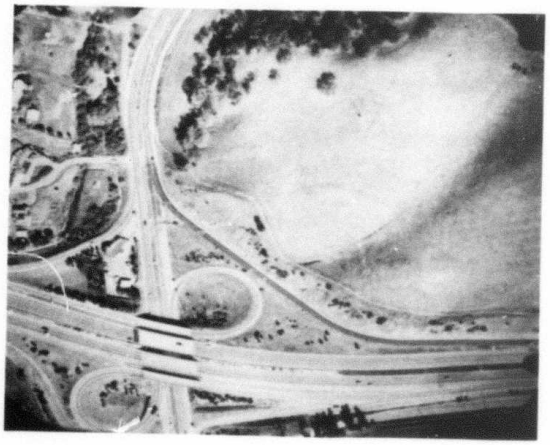


FIGURE 11 A TYPICAL AERIAL IMAGE TO BE CALIBRATED

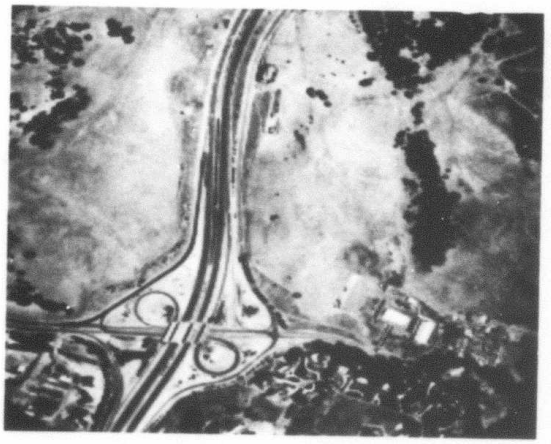


FIGURE 12 AN AERIAL IMAGE DISPLAYED SO THAT EACH PIXEL CORRESPONDS TO APPROXIMATELY 16 FEET ON THE GROUND

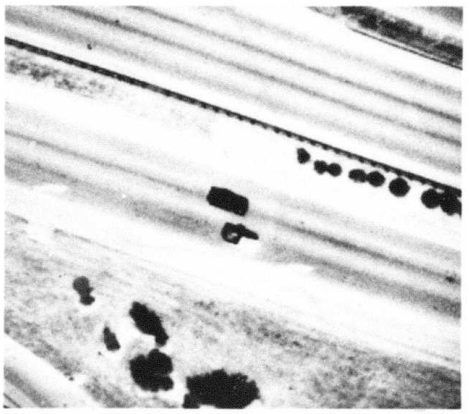


FIGURE 13 AN AERIAL IMAGE DISPLAYED SO THAT EACH PIXEL CORRESPONDS TO APPROXIMATELY 1 FOOT ON THE GROUND

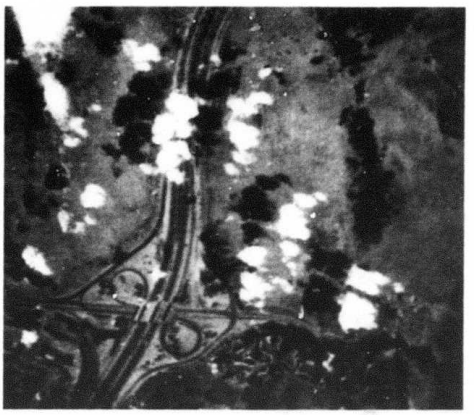


FIGURE 14 A TYPICAL IMAGE CONTAINING CLOUDS

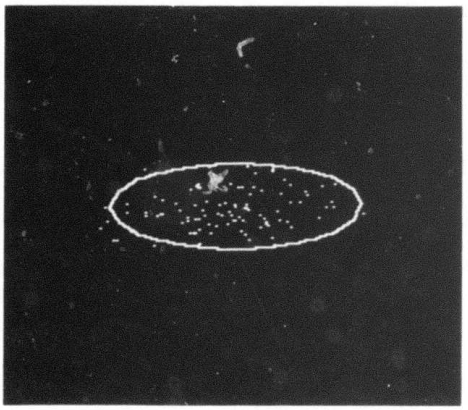


FIGURE 15 A PREDICTED UNCERTAINTY ELLIPSE AND A RANDOM DISTRIBUTION OF POSSIBLE LOCATIONS FOR THE FEATURE

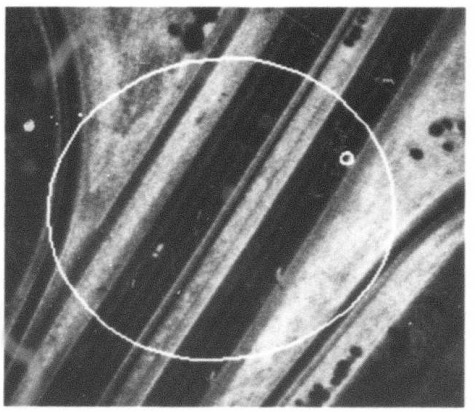


FIGURE 16 AN INITIAL UNCERTAINTY ELLIPSE AND A SMALL RELATIVE UNCERTAINTY ELLIPSE ABOUT A POINT FEATURE

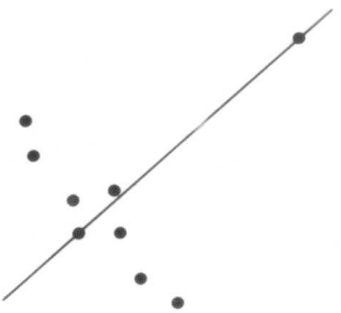


FIGURE 17 A PATHOLOGICAL EXAMPLE OF LEAST-SQUARES LINE FITTING

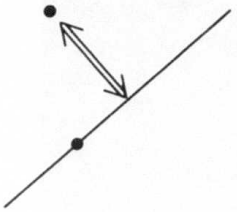


FIGURE 18(a) A POINT AND A LINE

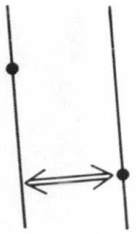


FIGURE 18(b) TWO PARALLEL LINES

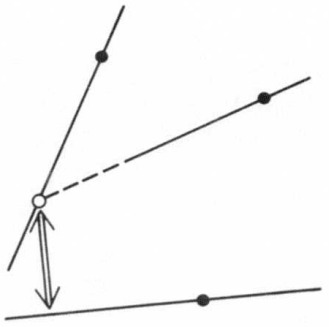


FIGURE 18(c) NON-PARALLEL LINES



FIGURE 19 A REFERENCE IMAGE OF THE SITE AND THE LOCATIONS OF THE POINT AND LINE FEATURES TO BE USED IN THE CALIBRATION

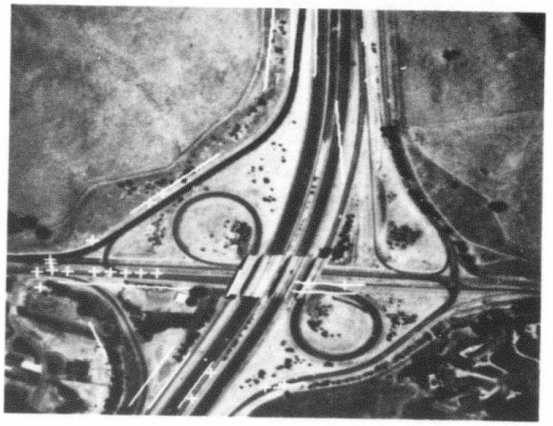


FIGURE 20 THE IMAGE TO BE CALIBRATED AND THE PREDICTED LOCATIONS OF THE FEATURES

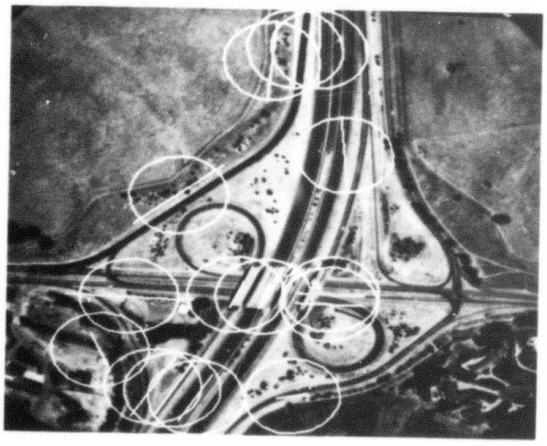


FIGURE 21 THE PREDICTED LOCATIONS OF THE ROAD SEGMENTS AND THE INITIAL UNCERTAINTY ELLIPSES ABOUT THEIR MID-POINTS



FIGURE 22 THE PREDICTED LOCATION OF A ROAD SEGMENT AND ITS MATCHING LOCATION



FIGURE 23 THE PREDICTED AND MATCHING LOCATIONS OF THREE ROAD SEGMENTS THAT ARE USED AS MUTUAL SUPPORT FOR EACH OTHER

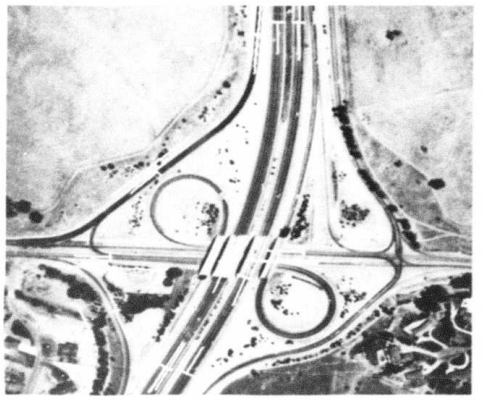


FIGURE 24 THE RESULTS OF ALL OF THE ROAD SEGMENT DETECTION OPERATORS

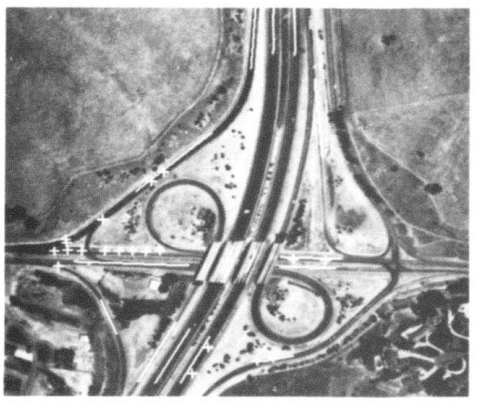


FIGURE 25 THE PREDICTED LOCATIONS OF THE FEATURES PRODUCED BY THE IMPROVED CAMERA PARAMETERS



FIGURE 26 AN IMAGE TO BE CALIBRATED THAT CONTAINS CLOUDS

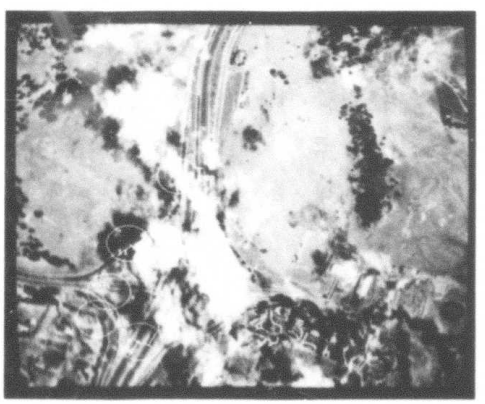


FIGURE 27 THE RESULTS OF ALL THE ROAD SEGMENT DETECTION OPERATORS

Appendix A

A LINEAR MODEL FOR PREDICTING THE DISTRIBUTION OF ERRORS UNDER A PROJECTIVE TRANSFORMATION

1. Problem Statement

GIVEN the set of camera parameters {yi} which define a projective transformation from 3-space to a 2-dimensional image plane {xi}, i=1,2; and assuming that the {yi}, i=1,2,...J, are jointly distributed according to a multivariate normal distribution function with given covariance matrix M, THEN we wish to find a region in the image plane, centered about the point provided by the projective transformation H{yi}, which will be large enough to contain the image of the corresponding 3-space point to some given level of probability.

2. Linear Approximation

As an approximation to the way in which the errors in the camera parameters produce displacements of a projected point, we will assume that:

$$\Delta x_1 = \sum_j \left(\frac{\delta x_1}{\delta y_j} * \Delta y_j \right)$$

[1] and

$$\Delta x_2 = \sum_j \left(\frac{\delta x_2}{\delta y_j} * \Delta y_j \right)$$

The partial derivatives in the above equations can be computed from the projective transformation H or measured experimentally. The two linear equations can be represented in matrix notation as:

[2] $$\Delta x = T(\Delta y)$$

where the transform T is the 2 x J matrix of the partial derivatives of the xi with respect to the yj, over the J camera parameters.

PRECEDING PAGE BLANK-NOT FILMED

To simplify our notation, we will assume that the image plane and 3-space coordinate axes have their origins at the projected and nominally imaged points respectively. Thus, the deltas in equation [2] can be dispensed with.

3. The Error Model

The multivariate normal probability density function has the form (for dimensionality "n"):

$$[3] \qquad P(X|U,M) = \frac{e^{(-.5 * (X-U)^T M^{-1} (X-U))}}{(2*\pi)^{\left(\frac{n}{2}\right)} * \sqrt{|M|}}$$

where: $U=E\{X\}$
$M=E\{(X-U)(X-U)^T\}$
$|A|$= determinant of A.

The covariance matrix M must be positive semidefinite. That is, for any n-dimensional vector Z with real components we have:

$$[4] \qquad Z^T MZ \geq 0.$$

Theorem 1[1]:

If Y is distributed according to [3] with mean vector U and covariance matrix M, then:

If $X=TY+B$ with T a constant matrix and B a constant vector, then X is normally distributed with mean $V=TU+B$ and covariance matrix $W=E[(X-V)(X-V)^T]=TMT^T$

Thus, given our previously stated assumptions, we can now assert that the error distribution in the image plane will be a bivariate normal probability density function, having the same form as equation [3], but with mean vector V, and covariance matrix W, obtained as described in the above theorem.

---

[1] T.W. Anderson, An Introduction to Multivariate Analysis, p. 25, (John Wiley & Sons, New York, New York, 1958).

In more explicit form we have:

$$[5] \qquad P(x_1,x_2 \mid 0,0,\rho,s_1,s_2) = \frac{e^{\left(-\frac{G}{2}\right)}}{2 * \pi * s_1 * s_2 * \sqrt{1-\rho^2}}$$

where:

$$G = \frac{\left(\frac{x_1^2}{s_1^2} - \frac{2 * \rho * x_1 * x_2}{s_1 * s_2} + \frac{x_2^2}{s_2^2}\right)}{(1 - \rho^2)}$$

$$s_1 = \sqrt{E\{x_1^2\}} \qquad s_2 = \sqrt{E\{x_2^2\}}$$

$$\rho = E\left\{\frac{x_1 * x_2}{s_1 * s_2}\right\}$$

We note that ρ is the coefficient of correlation between x1 and x2 and $(-1 \leq \rho \leq 1)$.

The contours of constant probability density in the image {x1,x2} plane are the loci where the exponent of the density function is constant. They are similar coaxial ellipses, with their axes parallel to the eigenvectors of the covariance matrix W. In particular, the major axis of the ellipse will make an angle of

$$[6] \qquad \alpha = \frac{1}{2} * \text{ARCTAN}\left(\frac{2 * \rho * s_1 * s_2}{\left(s_1^2 - s_2^2\right)}\right)$$

with the x1 axis.

To simplify our derivation of the dimensions of the ellipse needed to provide a given level of probability of containing the image of the 3-space point being projected, we will transform our coordinate axes in the image plane so that they lie along the major and minor axes of the coaxial constant probability ellipses.

The resulting covariance matrix Q has the form:

$$[7] \qquad Q = \begin{pmatrix} q_1^2 & 0 \\ 0 & q_2^2 \end{pmatrix}$$

where the qi (the new variances) are the eigenvalues of the covariance matrix W. These eigenvalues are found by solving the following equation:

$$[8] \qquad 0 = \begin{vmatrix} \left(s_1^2 - q^2\right) & \left(\rho * s_1 * s_2\right) \\ \left(\rho * s_1 * s_2\right) & \left(s_2^2 - q^2\right) \end{vmatrix} .$$

The resulting solutions are:

$$q_1^2 = \frac{1}{2} * \left(\left(s_1^2 + s_2^2\right) + \sqrt{\left(s_1^2 - s_2^2\right)^2 + 4 * \rho^2 * s_1^2 * s_2^2} \right)$$

[9] and

$$q_2^2 = \frac{1}{2} * \left(\left(s_1^2 + s_2^2\right) - \sqrt{\left(s_1^2 - s_2^2\right)^2 + 4 * \rho^2 * s_1^2 * s_2^2} \right)$$

Substituting $q1^2$ for q^2 in either of the two homogeneous equations in:

$$[10] \qquad 0 = \left(W - q^2 * I\right) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

allows us to solve for the ratio of the x1 to x2 coefficient in the major eigenvector and determine its angle with the x1 axis to be:

$$[11] \qquad \mathrm{TAN}(\alpha) = \frac{\left(q_1^2 - s_1^2\right)}{\left(\rho * s_1 * s_2\right)}$$

The above expression can be simplified using the identity ARCTAN(A)=2*ARCTAN({SQRT[1+A^2]-1}/A) to give the result in [6].

In terms of covariance matrix Q, the bivariate normal density function has the form:

$$[12] \qquad P(z_1,z_2) = \frac{e^{\left(-\frac{G}{2}\right)}}{2 * \pi * q_1 * q_2}$$

where: $$G = \frac{z_1^2}{q_1^2} + \frac{z_2^2}{q_2^2}.$$

The locus of $G=c^2$, where c is a constant is an equi-probability ellipse with major radius of length c*q1 and minor radius of length c*q2.

The area contained within this ellipse is c^2*q1*q2*PI and the differential area is 2*c*q1*q2*PI*Δc.

Thus, the probability p'' that the image of the nominally projected 3-space point will fall into the elliptic ring formed by the ellipses with parameters c and c+Δc is:

$$[13] \qquad P'' = c * e^{\frac{-c^2}{2}} * \Delta c.$$

Integrating p'' from 0 to c, we get:

$$[14] \qquad P = 1 - e^{-\frac{c^2}{2}}$$

where P is the probability that the image of the nominally projected 3-space point will fall into the ellipse with parameter c (i.e., the ellipse with major axis of length c*q1, minor radius of length c*q2, and orientation of the major axis of B; see equations [6] and [9] for the values of q1,q2, and α).

Some typical values for P are:

[15]

P	c
.50	1.177
.90	2.146
.95	2.448
.99	3.035

We note that if s1=s2=s, and =0, then q1=q2=s; the resulting contours are circles, and the parameter c corresponds to the radius of the resulting error circle measured in standard deviations (s). For this case, the radius which results in a 50% error probability is 1.177s, but the expected radial error is s*SQRT(PI/2)=1.253s, and the expected value of the square of the radial error is $E\{x1^2\}+E\{x2^2\} = 2*s^2$.

Finally, by invoking Bayes' theorem, we note that if an "error ellipse" as determined above is centered on the true projection of a given 3-space point, and has probability P of containing the actual projection of that point, then the same ellipse centered on the actual projection would have the same probability P of containing the true projection (assuming there is no difference in the way the true and actual projected points are distributed over the image plane).

Appendix B

RELATIVE UNCERTAINTY REGIONS

Let p and q be two three-dimensional feature points. Let a1 represent an estimate of the camera parameters. Let F represent the perspective transformation, which is a function of the camera parameters, that maps feature points into image points. Then

[1] $P = F(a1,p)$ and $Q = F(a1,q)$,

where P and Q are the two-dimensional image coordinates of the points p and q. P and Q are the predicted image locations for the two features based on the estimates a1.

If an operator has correctly located the image of p at P', where should the image of q be? Or, in which region should the image of q appear? That is, what is the relative uncertainty region about q with respect to p and P'?

Assume that the actual camera parameters are a2 and the two features actually appear at P' and Q' in the image. Thus,

[2] $P' = F(a2,p)$ and $Q' = F(a2,q)$.

The relative uncertainty region can be described by the difference between (Q' - P') and (Q - P) as a function of a1 and a2.

Let

[3] $$a2 = a1 + \Delta a .$$

If we make the same assumption made in Appendix A that the parameter space is locally linear about a1 and a2, then

[4] $$P' = F(a1,p) + Mp * \Delta a$$

and

[5] $$Q' = F(a1,q) + Mq * \Delta a$$

where Mp and Mq are the 2 x N matrices of partial derivatives that describe the relative changes in the image plane as a function of the N camera parameters. Then

[6] $[(Q' - P') - (Q - P)] = Mq * \Delta a - Mp * \Delta a$

or

[7] $[(Q' - P') - (Q - P)] = (Mq - Mp) * \Delta a.$

If the Δa's are distributed according to a multivariate normal distribution, Theorem 1 in Appendix A applies. If the mean of the distribution is the vector U and the covariance matrix is S, the vectors on the left side of linear equation [7] will be distributed with mean V = (Mq-Mp)*U and covariance matrix W = (Mq-Mp)*S*(Mq-Mp)T.

Appendix C

AN ITERATIVE METHOD TO REFINE CAMERA PARAMETERS

The standard calibration problem is:

> Assume that the correspondence between world points and image points is a perspective transformation, G, that is a function of several camera parameters, such as the X, Y, and Z position of the camera, the heading, pitch, and roll of the camera, and the focal length of the camera. Given an initial estimate of the camera parameters and a set of world points (Xi,Yi,Zi) and their corresponding image locations (Ui,Vi), determine the best (according to some error metric) camera parameter values to map the world points into the image points.

If G is a linear function of the camera parameters and the square of the unresolved errors is used as the metric, there is a standard solution to the problem. Let G be represented as a matrix M. Then for each world and image point pair:

[1]
$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix}.$$

A set of these equations can be combined into a single matrix:

[2]
$$\begin{pmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ \vdots \\ \vdots \end{pmatrix} = \begin{pmatrix} X_1 & Y_1 & Z_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_2 & Y_2 & Z_2 \\ & & & \ddots & & \\ & & & \vdots & & \end{pmatrix} \begin{pmatrix} M_{11} \\ M_{12} \\ M_{13} \\ M_{21} \\ M_{22} \\ M_{23} \end{pmatrix}.$$

Let A be the vector of U's and V's, P be the matrix of X's, Y's, and Z's, and B be the vector of M's. Then [2] can be restated as:

[3] $$A = P * B$$

This equation can be directly solved for the best least-squares solution for D, whose elements are the six elements of the matrix M[1]:

[4] $$B = (P^T * P)^{-1} * P^T * A$$

Unfortunately, G is generally not linear. However, the least-squares solution of the linear problem can be embedded in an iterative solution to a nonlinear problem. The idea is to approximate the surface about the estimated parameter values by a hyperplane, solve that linear problem, and iterate until the desired precision has been achieved. If the hyperplane is determined by the partial derivatives of G with respect to the camera parameters, this approach is similar to a multidimensional Newton-Raphson method. See [Gennery][2] or [Bolles][3] for a more detailed description of this approach.

In our calibration method we consider G to be a function of the following camera parameters:

Cx, Cy, Cz---the position of the camera
Ch, Cp, Cr---the heading, pitch, and roll of the camera
Cf---the focal length of the camera
Su, Sv---the image scale factors for the U and V directions
Ir---the image rotation about the piercing point
Du, Dv---the U and V position of the piercing point

We group them into two categories: "internal" camera parameters and "external" camera parameters. The idea is that the internal camera parameters are functions of the camera itself and generally remain constant from one picture to the next. They are the image scale factors, the image rotation, the piercing point location, and the focal length. The external camera parameters specify the position and orientation of the camera and generally vary from one picture to the

---

[1] F.A. Graybill, An Introduction to Linear Statistical Models, Vol. I, (Mc Graw-Hill Book Company, 1961).

[2] Donald B. Gennery, "Least-Squares Stereo-Camera Calibration," Stanford Artificial Intelligence Project internal memo (1975).

[3] Robert C. Bolles, "Verification Vision within a Programmable Assembly System," Stanford University Ph.D. Dissertation (December 1976).

next. Since the focal length may change in a camera that has a zoom lens, it is sometimes treated specially. It is separated out of the list of internal camera parameters and treated like an external camera parameter to be computed.

We use homogeneous matrices to represent the transformations that are functions of the parameters listed above. The internal or "digitization" matrix is defined to be:

$$[5] \quad D = \begin{pmatrix} S_u & 0 & 0 & 0 \\ 0 & S_v & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(I_R) & \sin(I_R) & 0 & 0 \\ -\sin(I_R) & \cos(I_R) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -D_u \\ 0 & 1 & 0 & -D_v \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We assume that D is constant and given as a priori information.

G is defined as follows:

$$[6] \quad G = D*F*R*P*H*T$$

where

$$[7] \quad F = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{C_F} & 0 \end{pmatrix} \qquad T = \begin{pmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & 1 & -C_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$H = \begin{pmatrix} \cos(C_h) & \sin(C_h) & 0 & 0 \\ -\sin(C_h) & \cos(C_h) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(C_p) & \sin(C_p) & 0 \\ 0 & -\sin(C_p) & \cos(C_p) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R = \begin{pmatrix} \cos(C_r) & 0 & \sin(C_r) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(C_r) & 0 & \cos(C_r) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

F is the perspective part of the camera transformation. T is the translation part. H, P, and R are the heading, pitch, and roll parts, respectively.

The transformation of world point (Xi,Yi,Zi) into an image point (Ui,Vi) is defined to be the following two-step computation:

$$[8] \qquad \begin{pmatrix} U_i' \\ V_i' \\ W_i' \\ S_i' \end{pmatrix} = D * F * R * P * H * \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}$$

$$U_i = \frac{U_i'}{S_i'} \qquad V_i = \frac{V_i'}{S_i'} .$$

In homogeneous coordinates S_i' is a scale factor for the vector and has to be divided out in order to obtain the image coordinates (Ui,Vi). Notice that Ui and Vi are not linear combinations of the camera parameters.

Given this representation of G, the partial derivative linear approximation to the surface in parameter space (about the initial estimates of the camera parameters) is:

$$[9] \qquad \begin{pmatrix} \Delta u_1 \\ \Delta v_1 \\ \Delta u_2 \\ \Delta v_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} \frac{\delta u_1}{\delta c_x} & \frac{\delta u_1}{\delta c_y} & \frac{\delta u_1}{\delta c_z} & \frac{\delta u_1}{\delta c_h} & \frac{\delta u_1}{\delta c_p} & \frac{\delta u_1}{\delta c_r} & \frac{\delta u_1}{\delta c_f} \\ \frac{\delta v_1}{\delta c_x} & \frac{\delta v_1}{\delta c_y} & \frac{\delta v_1}{\delta c_z} & \frac{\delta v_1}{\delta c_h} & \frac{\delta v_1}{\delta c_p} & \frac{\delta v_1}{\delta c_r} & \frac{\delta v_1}{\delta c_f} \\ \frac{\delta u_2}{\delta c_x} & \frac{\delta u_2}{\delta c_y} & \frac{\delta u_2}{\delta c_z} & \frac{\delta u_2}{\delta c_h} & \frac{\delta u_2}{\delta c_p} & \frac{\delta u_2}{\delta c_r} & \frac{\delta u_2}{\delta c_f} \\ \vdots & & & & & & \end{pmatrix} \begin{pmatrix} \Delta c_x \\ \Delta c_y \\ \Delta c_z \\ \Delta c_h \\ \Delta c_p \\ \Delta c_r \\ \Delta c_f \end{pmatrix} .$$

Since

$$[10] \qquad U_i = \frac{U_i}{S'_i} \text{ and } V_i = \frac{V'_i}{S'_i} ,$$

the partial derivatives have the form:

$$[11] \qquad \frac{\delta u_i}{\delta C_x} = \frac{S'_i * \frac{\delta u'_i}{\delta C_x} - U'_i * \frac{\delta s'_i}{\delta C_x}}{S'_i * S'_i} ,$$

which depends on the partial derivatives

$$[12] \qquad \frac{\delta u'_i}{\delta C_x} \text{ and } \frac{\delta S'_i}{\delta C_x} .$$

These partial derivatives can be computed as follows:

$$[13] \qquad \frac{\delta u'_i}{\delta C_x} = \frac{\delta}{\delta C_x} D * F * R * P * H * T * \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}.$$

And since most of the matrices are constants with respect to the variables being differentiated, these expressions can be greatly simplified. For example:

$$[14] \qquad \frac{\delta u'_i}{\delta C_x} = D * F * R * P * H * \frac{\delta T}{\delta C_x} * \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}.$$

In summary, the iterative method to refine camera parameters is to compute the partial derivatives shown above, form the linear

approximation shown in [9] for the error surface, use the method discussed for equation [2] to solve this linear problem for corrections to be added to the current estimate of the camera parameters, use the corrections to form new estimates of the camera parameters, and iterate this process until the unresolved errors are sufficiently small.

For point-on-a-line matches, instead of two constraints per match (i.e., Ui error and Vi error), only one constraint is added to the list of constraints accumulated in the matrix shown in [9]. That one constraint is based on the perpendicular distance between the point in the image that is supposed to be on the line and the predicted image of the line.

The distance between a point in the image, (Ui,Vi) and a line that passes through the point (u0,v0) and at angle θ with respect to the U axis is:

[15] $$d = (U_i - U_0) \sin\theta - (V_i - V_0) \cos\theta .$$

Therefore, the constraint for a point-on-a-line match adds one line to the partial derivative linear approximation:

[16] $$\Delta d_i = \left(\frac{\delta d_i}{\delta c_x} \frac{\delta d_i}{\delta c_y} \frac{\delta d_i}{\delta c_z} \frac{\delta d_i}{\delta c_h} \frac{\delta d_i}{\delta c_p} \frac{\delta d_i}{\delta c_r} \frac{\delta d_i}{\delta c_f}\right) * \left(\Delta c_x \Delta c_y \Delta c_z \Delta c_h \Delta c_p \Delta c_r \Delta$$

where each entry has the form:

[17] $$\frac{\delta d_i}{\delta c_x} = \left(\frac{\delta u_i}{\delta c_x} \sin\theta - \frac{\delta v_i}{\delta c_x} \cos\theta\right).$$

Each of these entries is a simple combination of the two partial derivatives used in the point-to-point case.

Notice that point-on-a-line matches and their constraints can be freely mixed with the normal point-to-point matches.

Appendix D

SYNTHESIS OF CLOUDS IN DATA BASE IMAGERY

In order to test the Road Expert under adverse viewing conditions, we considered it necessary to acquire images containing various degrees of cloud cover. Our primary source of imagery, CALTRANS (California Department of Transportation), does not photograph roads during cloudy weather conditions and therefore we had to synthesize the clouds appearing in our road images.

In order to generate realistic clouds in our test imagery, the following criteria were established:

(1) Clouds should cast shadows.

(2) Edges of clouds should be controllably wispy--no hard edges. The same should be true of cloud shadows.

(3) Interior of clouds should be controllably transparent.

Prototypical clouds were extracted from digitized 70 mm U-2 photographs by subtracting from each pixel a constant level CTHRESH which removed virtually all of the background while leaving the clouds intact.

The cloud prototype image was:

$$CLOUD[i,j] = MAX\ [\ (U2image[i,j] - CTHRESH),\ 0\ ]$$

The following ramp function was introduced to satisfy b):

$$RAMP[i,j] = MIN\ [\ (CLOUD[i,j]/RAMPLEVEL),\ 1]$$

The ramp function assumes that cloud edges and partially transparent interiors of clouds have photometric levels close to zero (CTHRESH in the U-2 image). The "width" of the ramp is set indirectly by the selection of the intensity level "RAMPLEVEL."

Shadows are introduced by:

```
SHADOWIMAGE[i,j] = ROADIMAGE[i,j]
                   * (1 - (1 - GROUND.ATTEN) * RAMP[i+di,j+dj])
```

where di,dj define to offset of the cloud shadow with respect to the cloud. Clouds are assumed to be at a constant height above the underlying terrain. It is easily seen that when RAMP[i,j]=0, the image is unaffected, and when RAMP[i,j]=1, the image is attenuated by factor GROUND.ATTEN. Clouds are introduced to the shadow image by:

```
CLOUDIMAGE[i,j] = SHADOWIMAGE[i,j] * (1 - RAMP[i,j])
          + RAMP[i,j] * (CLOUD[i,j] * CLOUD.CONTRAST.FACTOR
                                    + CLOUD.INTENSITY.OFFSET)
```

This function smoothly blends the clouds with the shadowed road image according to the same ramp function.

The above procedure for synthesizing clouds has a total of seven parameters which control the attenuation of the ground intensity due to the cloud shadows and the clouds; control the blending at the cloud edges; control the relative contrasts of the clouds with respect to the ground; and finally, set the spatially offset of the cloud shadows with respect to the clouds.

Annex C

Detection of Roads and Linear Structures in Aerial Imagery by Computer

(From SRI Technical Note 200, October 1979;
also in Proceedings Image Understanding
Workshop, November 1979, pp. 87-100)

ABSTRACT

This paper describes a computer-based approach to the problem of detecting and precisely delineating roads, and similar "line-like" structures, appearing in low-resolution aerial imagery. The approach is based on a new paradigm for combining local information from multiple, and possibly incommensurate, sources, including various line and edge detection operators, map knowledge about the likely path of roads through an image, and generic knowledge about roads (e.g., connectivity, curvature, and width constraints). The final interpretation of the scene is achieved by using either a graph search or dynamic programming technique to optimize a global figure of merit. Implementation details and experimental results are included.

I INTRODUCTION

A person given the problem of producing an overlay showing the clearly visible roads in an aerial image would normally be expected to accomplish this task with little difficulty, even though he may be completely unfamiliar with the terrain depicted in the image. Our purpose in this paper is to clarify the nature of this task and some of its generalizations. In particular, we wish to specify the requirements and mechanisms for a machine to be capable of near-human performance in finding roads and other semantically meaningful linear structures in aerial images.

A. Performance Criteria

Our goal is to produce a list of connected points for each segment of road which is tracked in the input image. Each such track is a delineation of the actual road and should have the following properties:

(1) No point on a track should be located outside of the road boundaries when the road is clearly visible.

(2) The track should be smooth where the road is straight or smoothly curving (within the constraints of a digital raster representation).

(3) If parts of the road are occluded, those portions of the continuous track overlaying the occluded segments should be labeled as such.

(4) In areas where the road is partially occluded, the track should follow the actual center of the road (as opposed to the center of the visible portion). If the road is composed of adjacent but separated lanes, then each lane will be considered a separate road for our purposes.

B. Contextual Settings for Road Tracking

A "road" is a functionally defined entity whose appearance in an image depends largely on its width and how much internal road detail is visible; i.e., appearance depends largely on image resolution (see Figure 1: Road Scenes Depicted At A Spectrum of Resolutions).

Additional factors having a major effect on visually locating roads in imagery include the visible extent of the road, its contrast with the adjacent terrain, the presence of nearby linear structures, and any prior knowledge about the actual shape of the road and its location in the image.

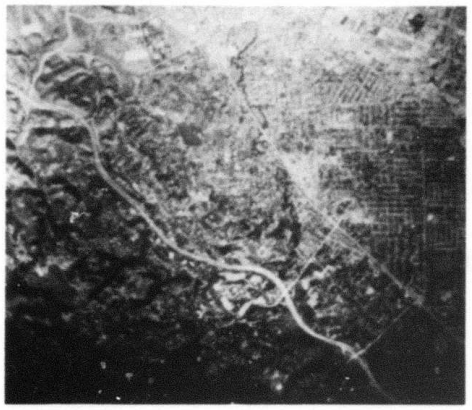
We have found that the following contextual settings require significantly different approaches to the road tracking problem:

(1) High vs. low resolution (low resolution is defined as the case in which the road being tracked has an image width of three or fewer pixels).

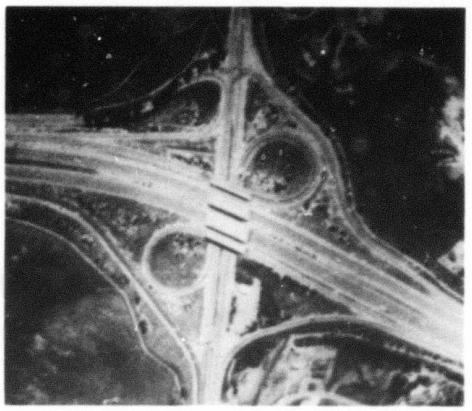
(2) Clear vs. occluded viewing (clear viewing is defined as a situation in which no more than approximately 30% of the road being tracked is occluded by clouds, intervening objects, etc.).

(3) High vs. low density of linear detail (nominally, this distinction corresponds to urban vs. rural scenes).

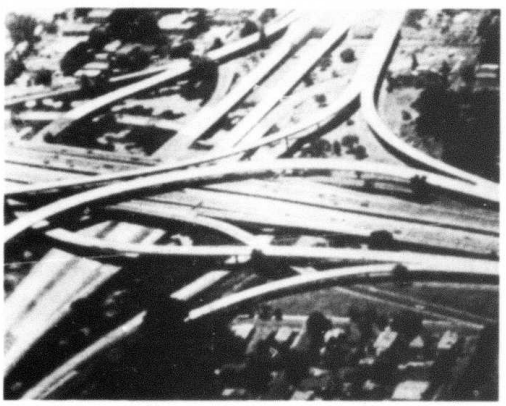
In this paper we will mainly be concerned with tracking roads in clear imagery of rural scenes at low resolution. A robust technique for tracking roads in high-resolution imagery was previously reported (Quam [1978]). We note that in the case of high-resolution imagery, once the road has been "acquired" and we are able to track features internal to the road boundaries, the surrounding detail is of minor importance (except as it introduces shadows and occlusions); thus, the distinction between urban and rural scenes is important mainly at low resolution. Where the roads are heavily occluded, road matching rather than road tracking is the appropriate technique; here one needs to have prior knowledge of the geometry of the road networks being searched for. Prior knowledge about the (approximate) location and/or direction of the roads in the imagery is important if a specific road (as opposed to all roads) is to be tracked; some method of indicating which road we are interested in is necessary, and this is typically done by delimiting a search area in the input image. Finally, prior knowledge about terrain type and/or scene elevations can be used to help distinguish low-resolution roads from other linear features by invoking cultural and economic constraints which are known to affect road construction.



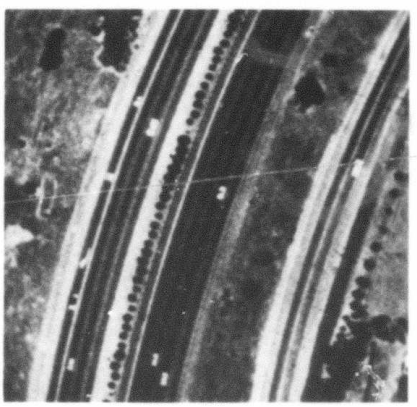
(a)



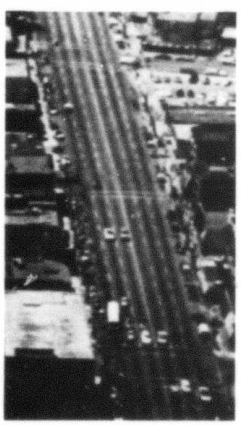
(b)



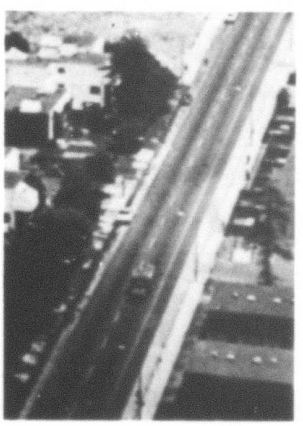
(c)



(d)



(e)



(f)

FIGURE 1 ROAD SCENES DEPICTED AT A SPECTRUM OF RESOLUTIONS

II LOW-RESOLUTION ROAD TRACKING

At low resolution roads are often indistinguishable from other linear features appearing in the image (including artifacts, such as scratches). Thus, the low-resolution road tracking problem largely reduces to the general problem of line (as opposed to edge) following. Nevertheless, there are still some weak semantics that can be invoked to specifically tailor a system for road tracking, trading some generality for significant increases in performance.

A. The Basic Paradigm

The basic paradigm we employ is to first evaluate all local evidence for the presence of a road at every location in the search area (a low numeric value indicates a high likelihood that the given image point lies on a road), and then find a single track which, while satisfying imposed constraints (such as continuity), minimizes the sum of the local evaluation scores (costs) associated with every point along the track. While the basic optimization paradigm is not new (e.g., Fischler [1973], Montanari [1971], Martelli [1976], Barrow and Tenenbaum [1975] Rubin [1978]), it is incomplete in that it does not provide mechanisms for reconciling incommensurate sources of information. This capability is crucial in problems such as road tracing in which no single coherent model is adequate for reliable detection. In this paper we introduce new and relatively simple mechanisms for combining local evidence and constraints in the context of an optimization paradigm for detecting linear structures.

B. Detecting Local Road Presence--Road Operators and Models

At low resolution roads are line-like structures of essentially constant width, which, in general, are locally constant in intensity in the along-track direction and show significant contrast with the adjacent terrain (generally, they are either uniformly lighter or darker). A specific interpretation of this low-resolution road model is

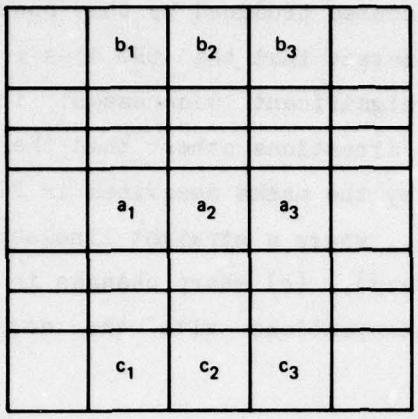
embodied in the Duda Road Operator (DRO)* described in Figure 2. In Figure 3 we show some examples of the scores produced by this operator on a variety of road scenes. It is apparent that the DRO does a good job most of the time but has some significant weaknesses; it is sensitive to (a) road orientation (in directions other than the four principal directions explicitly covered by the masks described in Figure 2), (b) raster quantization effects (e.g., where a straight line segment "jogs" in crossing a quantization boundary), (c) sharp changes in road direction, and (d) to certain contrast problems with the adjacent terrain.

At this point one might wonder if a special road operator is really required; why not simply use a generic edge detector (e.g., Sobel [in Duda and Hart, 1973], Roberts [1965], or Heuckel [1971 and 1973])? Even more to the point, we notice that it is possible to interpret the effect of employing an operator on an image as resulting in the suppression of all detail other than that associated with the entity to be detected; therefore, a high-pass filter might act as a perfectly good road operator. Finally, roads will generally be lighter or darker than the immediately adjacent terrain; why not simply use the actual intensity values (contrast-enhanced and possibly inverted, depending on the relative brightness between the road and adjacent terrain)? In Figure 4 we show a comparison of these different techniques applied to the same road scene; in Figure 5 the scores are thresholded to make explicit the locations in the image which are assigned the highest road presence likelihoods by the different techniques.

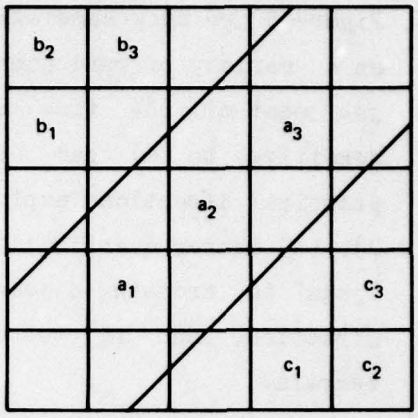
In the approach we have developed, a key attribute characterizing the utility of a "local" image feature detector (i.e., "operator") is the percentage and coherence of its mistakes when it is almost certain it has found instances of the feature it is designed to detect. Even though the Duda road operator makes mistakes of omission, its performance in not making coherent false-alarm type errors is quite good.

---

*Suggested by R. O. Duda of SRI International.



MASK TO DETECT HORIZONTAL ROAD SEGMENTS

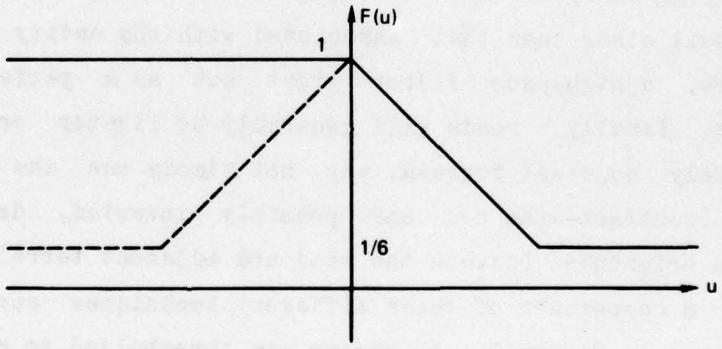


MASK TO DETECT RIGHT DIAGONAL ROAD SEGMENTS

$$\text{SCORE: } 100 - [G(a_1 - a_2) \times G(a_2 - a_3) / \sum_{i=1}^{3} F(a_i - b_i) + F(a_i - c_i)]$$

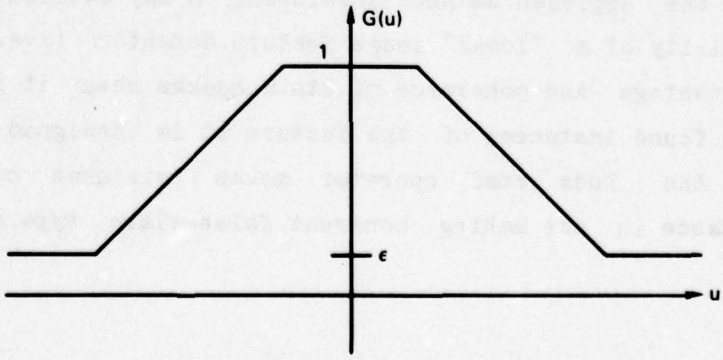
WINDOWS FOR THE ROAD OPERATORS

Two other masks, similar to these, are used to detect vertical and left diagonal road segments.



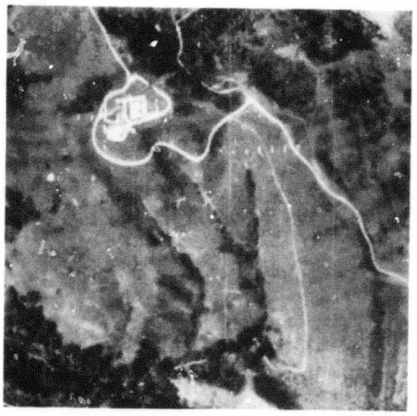
ROAD EDGE SCORING FUNCTION

Function depicted as solid line is used for light roads. F(-u) is used for dark roads. Symmetric form of the function, shown by dashed lines for negative values of u, is used when road to background contrast is unknown.

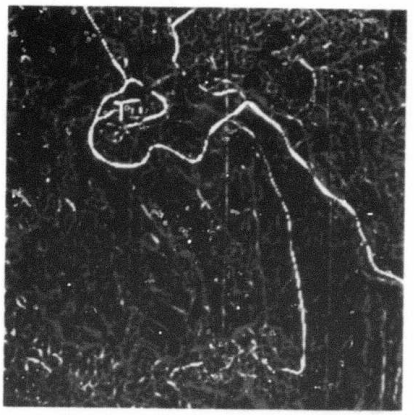


ROAD UNIFORMITY SCORING FUNCTION

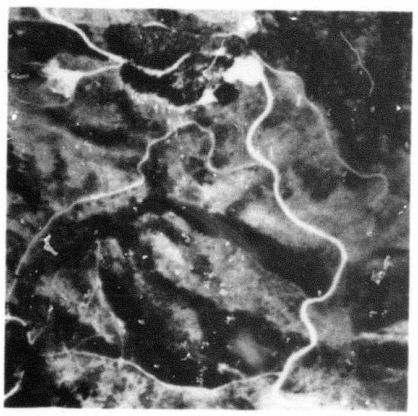
FIGURE 2 DESCRIPTION OF DUDA ROAD OPERATOR



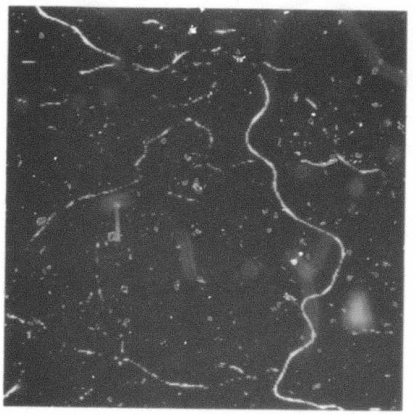
(a)



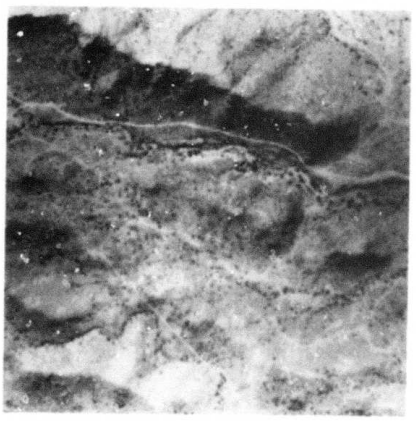
(b)



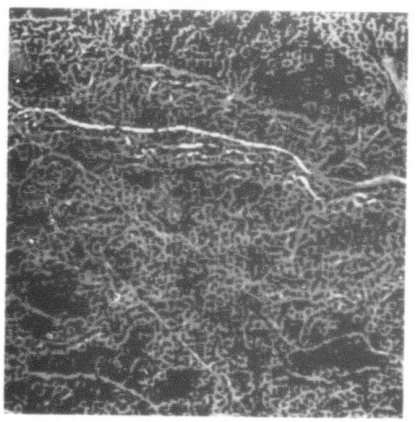
(c)



(d)

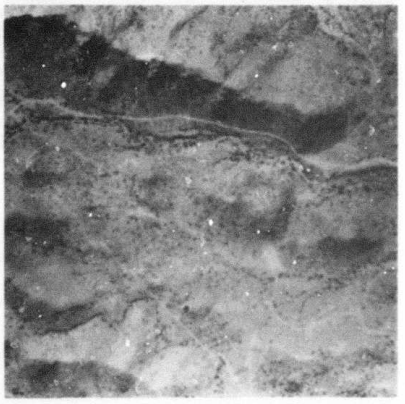


(e)

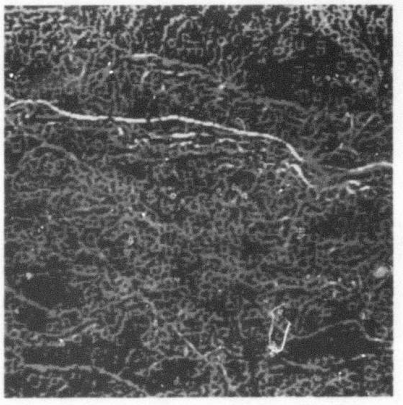


(f)

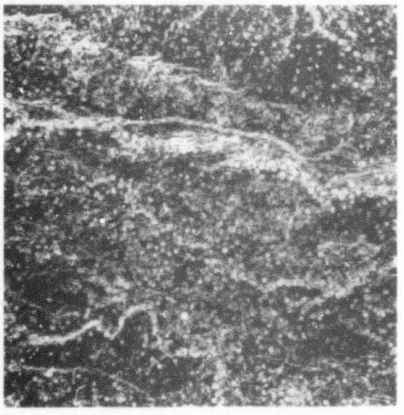
FIGURE 3 DUDA ROAD OPERATOR APPLIED TO A NUMBER OF SCENES



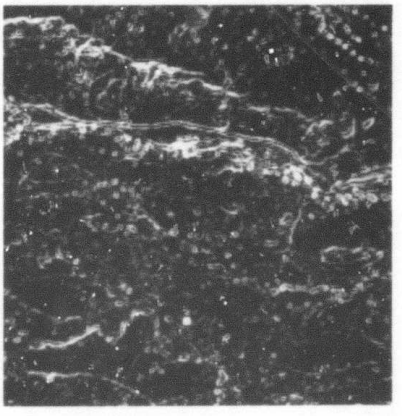
(a) ORIGINAL IMAGE



(b) DUDA ROAD OPERATOR



(c) ROBERT'S CROSS GRADIENT



(d) SOBEL-TYPE GRADIENT

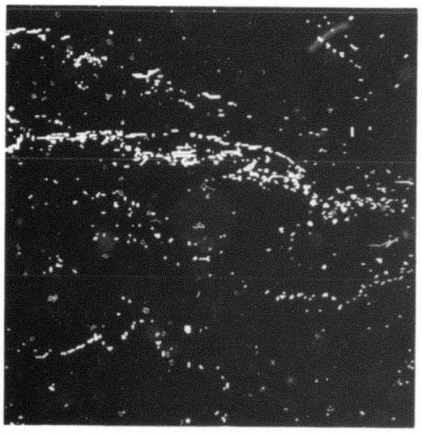
FIGURE 4 DIFFERENT ROAD OPERATORS APPLIED TO THE SAME SCENE



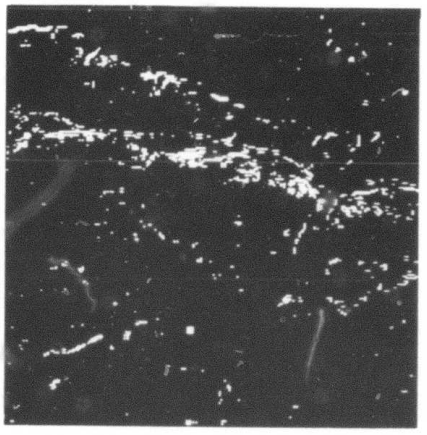
(a) ORIGINAL IMAGE



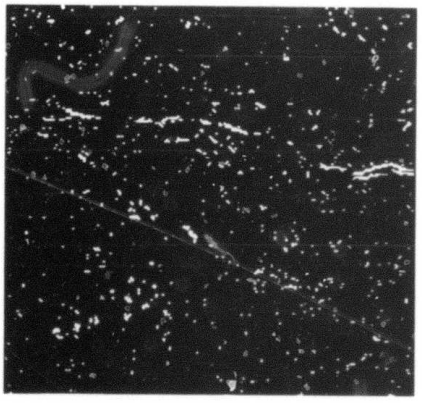
(b) DUDA ROAD OPERATOR



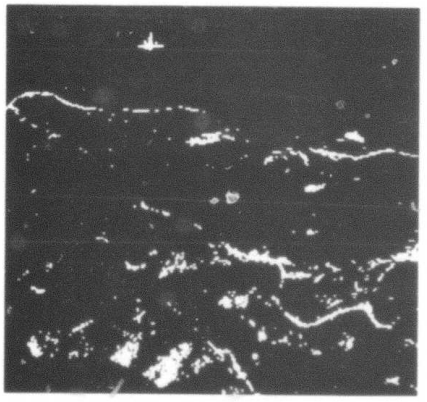
(c) ROBERT'S CROSS GRADIENT



(d) SOBEL-TYPE GRADIENT



(e) HUECKEL LINE OPERATOR



(f) INTENSITY

FIGURE 5 DIFFERENT ROAD OPERATORS APPLIED TO THE SAME SCENE
(Operator scores are thresholded to highlight the locations assigned the best scores.)

C. Combining Incommensurate Sources of Knowledge--An Elaboration of the Basic Optimization Paradigm

We will now specify a general approach for combining the results deduced by the application of a set of (road) operators, as well as to the problem of allowing prior knowledge and constraints to influence the answer produced by the optimization algorithm.

We partition our inventory of operators into two categories--Type I operators, each of which can be adjusted to make very few coherent errors in detecting instances of the relevant feature when the feature is not present (possibly at the cost of making a large number of omission errors); and Type II operators, each of which can be adjusted to reliably give a quantitative indication of the presences of the feature when it is actually under examination (but these operators might be very unreliable in their assertions when examining something other than the desired feature). Our basic approach is to strongly bias (or even constrain) the desired answer to fit the coherent pattern produced by a superposition of evidence provided by all the Type I operators and to fill in the details locally, using that particular Type II operator which seems to be most certain that it has found the desired feature. (A more comprehensive discussion on methods for combining multisource evidence is given in Fischler and Garvey [in preparation].)

A problem that immediately arises is how to combine the results of several Type I and Type II operators. By considering the output of Type I operators to be valid binary decisions, we have made them commensurate and can logically combine their outputs. In the context of tracking roads (or other linear features), we scan each of our Type I operators over some specified region of interest and create a binary overlay mask containing the logical union of the locations at which one or more of these operators has detected (with high likelihood) the presence of a road. An example of such a mask, called a "perfect road score" (PRS) mask, is compared in Figure 6 with the road image from which it was derived.

The problem of combining the results produced by a set of Type II operators has no acceptable solution when the values they return are not probabilities nor other commensurate quantities. However, Type I and Type II operator scores can be combined, since a positive Type I output can always be set to the maximum value (zero cost) on the likelihood scale of any Type II operator.

Our approach is thus to AND the PRS mask (containing the union of all positive Type I outputs) with every array of scores produced by the Type II operators to produce a set of cost arrays (CA) with zero cost scores at the locations marked in the PRS mask. The optimization algorithm is separately applied to each CA, and the path with the lowest global score is selected as the primary road track through the given region.

In addition to creating a framework for "growing" the road using the Type I operators, we have developed a simple mechanism for introducing constraints and a priori information via the scores obtained from the Type II operators. This is accomplished by numerically transforming the value "x" originally produced by any Type II operator using the function: score = x↑a+b (with control constants a and b). For example, if control constant "a" is held fixed and "b" is increased, the resultant optimal path through the CA would tend to be smoother and straighter (somewhat like pulling the path taut); this effect occurs because, as "b" is increased, the length of the path becomes relatively more important in comparison to the local quality as defined by the individual values "x" returned by the operator. If we are tracking a rocky coastline in an image, we would opt for placing the path through the locations having the best edge scores as opposed to trying to smooth the result; here we would use a zero value for "b". In the case of tracking a road where smoothness is a nominal property, we would select some nonzero value for "b". If we had a priori information that a road we are attempting to track is fairly straight, we could use a high value for "b".

As the value of control constant "a" is increased, there is a very strong inhibition against going through a point having a low likelihood of being on a road. Thus, if we wish to track a road in a region where there may be other strong linear structures nearby, a high value of constant "a" will prevent a jump from one linear object to another; but this can result in wandering (e.g., around shadows, vehicles, etc., in the case of tracking a high-resolution road). Figure 7 shows some examples of tracking a road with different values of the two control constants.

D. The Low-Resolution Road Tracking Algorithm (LRRT)

The low-resolution road tracking algorithm operates as follows:

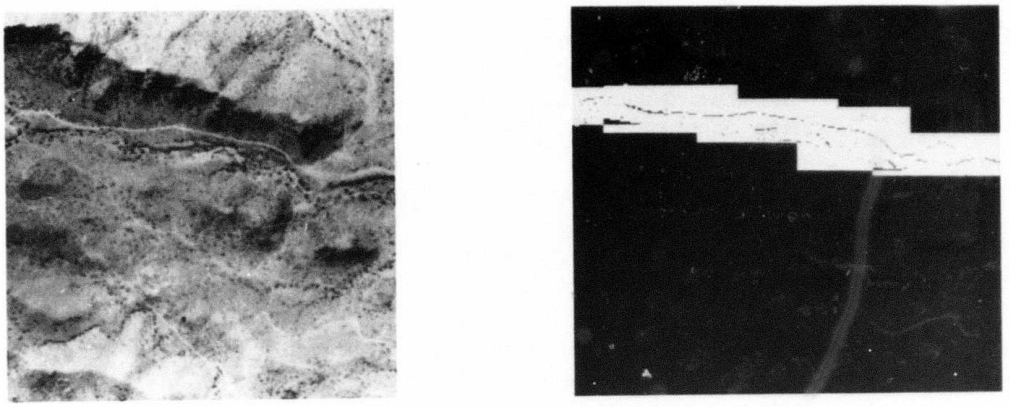
(1) A search region is designated in the image. This search region is defined by a binary mask which delimits the search for the road track.

(2) A selected set of Type I operators are scanned over the region designated by the search mask; and the scores produced by each such operator are histogrammed and thresholded at some preset level, based both on the nature of the operator and so that the number of points below this threshold will not exceed the number of road points estimated to be present in the search window (e.g., selecting 5% of the points in the search window might be an upper limit for the Duda road operator). A PRS mask is generated as the union of those locations at which each Type I operator is lower than its associated threshold (scores are treated as costs; a lower score implies a more road-like appearance).

(3) A selected set of Type II operators is scanned over the region designated by the search mask, and the scores produced by each such operator are either scaled or normalized (e.g., by their histogram ranking) to lie in the nominal range from 1-100; the scores for each Type II operator are stored in a separate array.

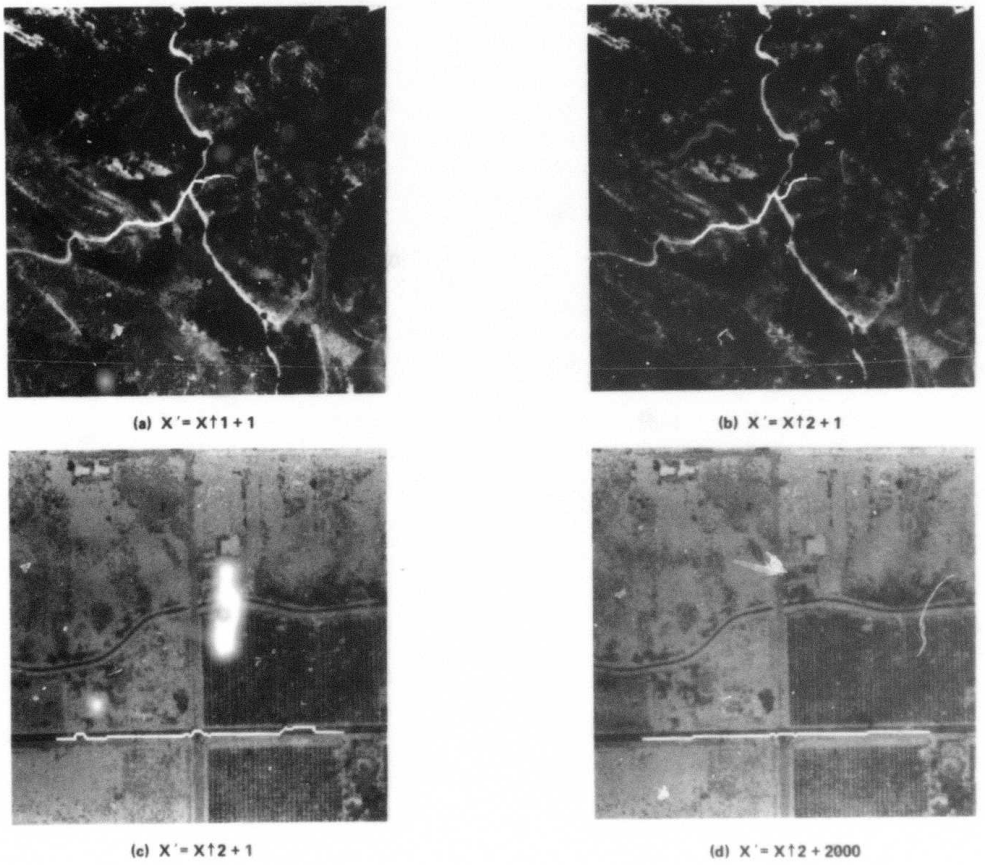
(4) Each Type II array is now sequentially modified as indicated:

   (a) In those regions of the image where some external source of information indicates that occlusions exist (e.g., due to clouds or to intervening objects), or where there is no significant contrast between the road and the adjacent terrain, thus rendering the road



(a) (b)

FIGURE 6 A SCENE AND ITS PERFECT ROAD SCORE MASK



(a) $X' = X\uparrow 1 + 1$ (b) $X' = X\uparrow 2 + 1$

(c) $X' = X\uparrow 2 + 1$ (d) $X' = X\uparrow 2 + 2000$

FIGURE 7 EXAMPLES OF HOW TRANSFORMING TYPE II IMAGE OPERATOR SCORES (X) ALLOW US TO ADJUST THE TRADE-OFF BETWEEN ROAD SMOOTHNESS AND PLACING THE ROAD TRACK AT ITS LOCALLY MOST PROBABLE LOCATION

"invisible," a constant is added to the score at each pixel location. This is done in order to reduce the preference for one path over another through areas where the local operators are incapable of returning relevant information about road presence.

(b) The scores at those locations corresponding to points in the PRM are set to zero (actually, they are set to some very small positive value to prevent arbitrary wandering, or even cycling, through regions of zero cost).

(c) Every score "x" in the array is transformed (as discussed earlier) by the formula:

$$x' = x\uparrow a+b \quad .$$

This transformation allows us to introduce external information in adjusting the balance between track smoothness (or straightness) and placing the track at its locally most probable location.

(5) Starting and terminating delimiters are designated in the search area: either a pair of lines (e.g., the sides of the search window) or a pair of boxes, through which the road must pass. Each Type II cost array is considered to be a graph with each pixel connected to each of its eight neighbors, and a minimum cost path is found in each such array between the starting and terminating delimiters. Since there is no way to directly compare the relative merits of road tracks computed in two distinct Type II arrays, we employ a heuristic in which the average score per pixel along the track in each Type II array is computed, and its histogram ranking in comparison with all the scores produced by the given Type II operator over the search window is taken as the quality of the track. The track with the highest quality number is chosen as the preferred track.

III THE GENERAL PROBLEM OF LOW-RESOLUTION ROAD TRACKING (MULTIPLE ROADS)

We find that it is desirable to deal with the road-(linear feature)-tracking problem in three distinct phases:

(1) The first phase produces a crude delineation of all the roads to be tracked (either producing an approximate track for each road segment or narrowing the search areas containing the different road segments). This delineation can be obtained by making multiple passes through the initial search area of an image with the LRRT described above. After each pass, the detected road track is marked as a forbidden area so as to allow the *next most prominent* road segment to be detected. If two distinct road tracks have common segments or have the same start and stop delimiters, then the "suboptimal" road tracks produced by the linking algorithm (the algorithm which finds the lowest-cost path through the Type II operator cost array) will generally delineate additional road segments.

With the availability of an external knowledge source, such as a map data base or a sketch map, the desired delineation can be obtained more directly.

(2) The second phase produces a precise track for each road segment of interest by applying the LRRT to the individual crude delineations obtained in the first phase.

(3) The third phase involves smoothing and possibly linking road segments separated by regions of significant occlusion, as well as marking those portions of a road track that were inferred from continuity rather than direct visibility.

IV IMPLEMENTATION DETAILS AND EXPERIMENTAL RESULTS

While we have addressed the problems associated with each of the above phases for automatically delineating the low-resolution roads and linear structures in an image, most of our current experimental work has been concerned with obtaining a high-performance solution to the problem of precise delineation required in phase two. We have implemented two

versions of the LRRT generically described in the preceding section: an INTERLISP/SAIL version for developmental work and a FORTRAN version for more extensive experimentation and evaluation. Both versions run on the SRI PDP-10, while the FORTRAN version was designed to also be compatible with a CDC 6400 system at the U. S. Army Engineer Topographic Lab (ETL) at Ft. Belvoir (the FORTRAN version has a minimum core requirement of 20,000 60-bit words, and will track a road segment 128 pixels long in 15 seconds of CPU time; the corresponding numbers for the INTERLISP version are 90,000 36-bit words of core and 60 seconds of CPU time).

The FORTRAN version of the LRRT makes some additional assumptions about the roads to be tracked: it assumes that they are generally lighter or darker than the surrounding terrain and that they do not "double back" on themselves in the designated search areas. It uses a single Type II operator (based on histogram normalized image intensity) and two Type I operators (the Duda road operator and an image intensity operator, which thresholds image intensity and also checks the size of the above threshold intensity region about a potential road point to determine if the width constraint is satisfied). This program has already been tested on approximately fifty road segments found in aerial images of seven different geographic locations with no failures, where the assumptions are satisfied and the roads are clearly visible (some examples are shown in Figure 8).

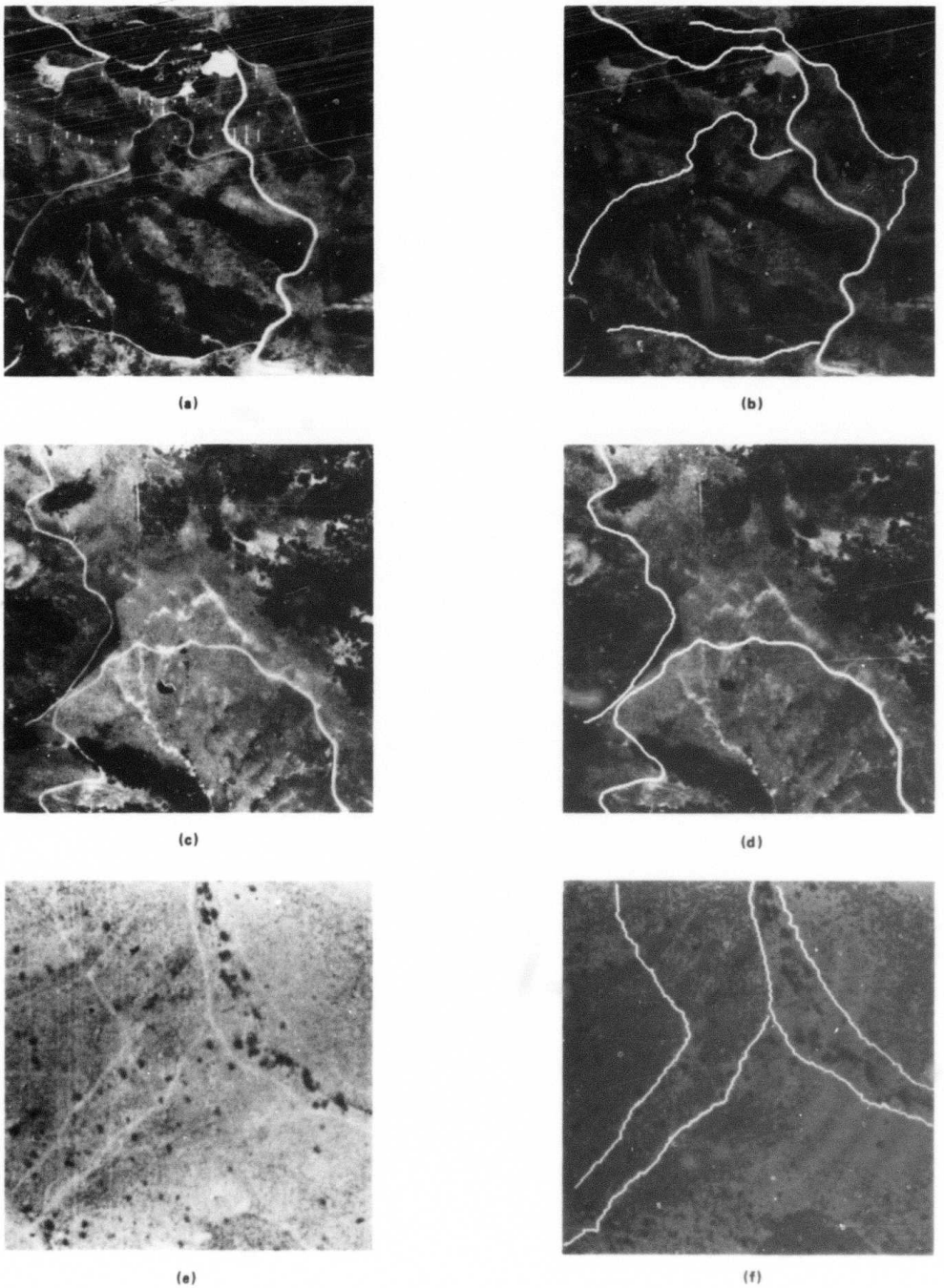
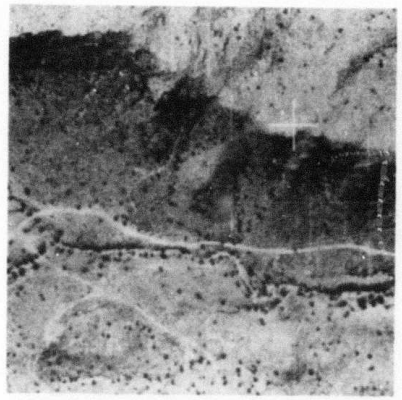
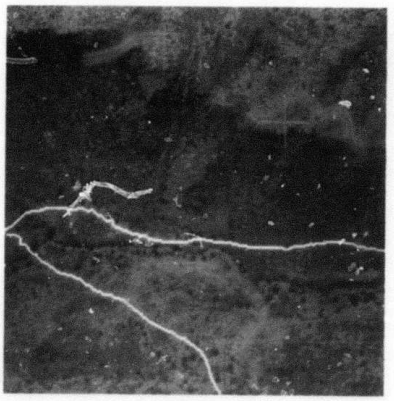


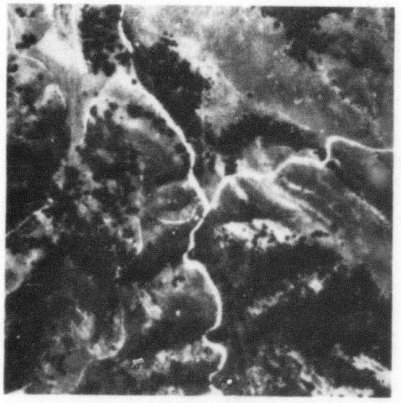
FIGURE 8 EXAMPLES OF ROAD DELINEATION PRODUCED BY THE LOW RESOLUTION ROAD TRACKER



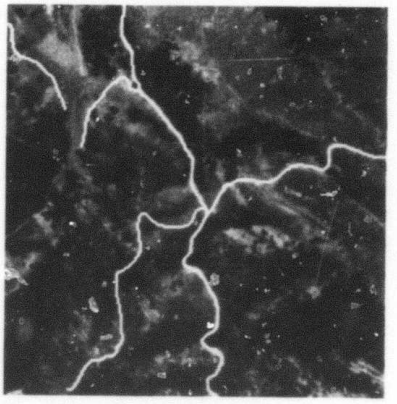
(g)



(h)



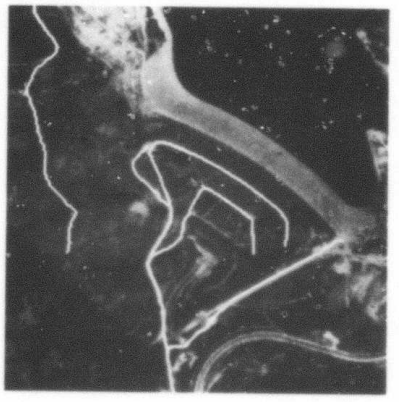
(i)



(j)



(k)



(l)

FIGURE 8 EXAMPLES OF ROAD DELINEATION PRODUCED BY THE LOW RESOLUTION ROAD TRACKER (Concluded)

V CONCLUDING COMMENTS

In this paper we have addressed the problem of precise delineation of the roads and linear features appearing in aerial photographs using an approach based on global optimization of locally evaluated evidence. Since there does not appear to exist a single coherent model suitable for reliable detection of local road presence, it was essential that some means for integrating information from multiple (incommensurate) image operators and knowledge sources be devised--the conventional optimization paradigm does not provide any formal machinery for achieving this task.

Two key points characterize the basis of our approach:

(1) Rather than projecting all image operators on a single linear scale and attempting to use them in the same qualitative manner, we have identified the distinctly different nature and potential use of operators which have strong object detection capabilities as opposed to those which are useful for object analysis once identification and/or location is known. (Depending on the specific context, a particular operator might switch from one role to the other.) We have provided a simple and uniform mechanism for integrating the information provided by the two classes of operators for the specific task of tracking linear structures, and we believe that the same general approach is applicable in a wider range of problem settings.

(2) We have recognized the fact that the score returned by an image operator usually has little absolute meaning, and yet a monotonic transformation of this score can lead to a significantly different final result in tracking linear structures. We have capitalized on this property by introducing a monotonic transform which allows a simple and uniform mechanism for adjusting the scores to reflect a priori information and semantic constraints.

Our plans for future work include the development of more effective techniques for the completely unconstrained delineation required in phase one (defined earlier), for tracking and possibly distinguishing among a variety of different types of linear structures (e.g., roads, rivers, railroads, runways, etc.), and for tracking linear structures in three dimensions using stereo image pairs.

The scientific content of this work lies in discovering effective models for representing and detecting the linear structures of interest and developing paradigms for integrating information from the wide variety of knowledge sources available to the human observer whose performance we are attempting to equal or surpass. Applications of our work in the military area include road monitoring for intelligence purposes, delineation of roads and linear features for automated cartography, and detection of roads and linear features as landmarks for autonomous navigation and weapon guidance.

REFERENCES

1. M. A. Fischler and R. A. Elschlager, "The Representation and Matching of Pictorial Structures," IEEE Trans. on Computers, Vol. C-22, No. 1, pp. 67-92 (January 1973).

2. M. Heuckel, "A Local Visual Operator which Recognizes Edges and Lines," J. ACM, Vol. 20, No. 4 (October 1973).

3. M. Heuckel, "An Operator which Locates Edges in Digitized Pictures," J. Assoc. Computing Mach., Vol. 18, pp. 113-125 (1971).

4. A. Martelli, "An Application of Heuristic Search Methods to Edge and Curve Detection," Comm. ACM, Vol. 19, No. 2 (February 1976).

5. U. Montanari, "On Optimal Detection of Lines in Noisy Pictures," Comm. ACM, Vol. 14, No. 5 (May 1971).

6. L. G. Roberts, "Machine Perception of Three-Dimensional Solids," Optical and Electro-Optical Information Processing, Tippett et al., eds., pp. 159-197 (M.I.T. Press, Cambridge, Massachusetts, 1965).

7. L. Quam, "Road Tracking and Anomaly Detection," in Proceedings: Image Understanding Workshop, pp. 51-55 (May 1978).

8. S. Rubin, "The ARGOS Image Understanding System," Ph.D. Thesis, Dept. of Computer Science, Carnegie-Mellon University, Pittsburgh, Pennsylvania (1978).

9. H. G. Barrow and J. M. Tenenbaum, "The Representation and Use of Knowledge in Vision," A.I. Technical Note 108, SRI International, Menlo Park, California (July 1975).

10. R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis (Wiley, 1973).

11. M. A. Fischler and T. D. Garvey, "A Computer-Based Approach to Perceptual Reasoning and Multisensor Integration" (in preparation).

Annex D

Road Tracking and Anomaly Detection

(From "Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, SRI Project 5300, May 1978, pp. 13-23)

II ROAD TRACKING AND ANOMALY DETECTION

This section of the report describes a new procedure for tracking road segments and detecting potential vehicles in aerial imagery of approximately 1 to 3 feet per pixel ground resolution. The road tracking algorithm discussed here is currently initiated by manually specifying the position of the center and the direction of the road fragment we wish to analyze. The nominal road width could be supplied by the user, by the data base, or by an image analysis function that can determine the width of a road fragment. The road tracker produces two forms of output: a point list describing the track of the road center, and a binary image of all points in the road that are anomalous and might represent to vehicles. In the complete road-expert system, this image will then be analyzed to screen false alarms and interpret the remaining anomalies.

A. Algorithm Description

Figure 14 shows a representative road scene digitized at a ground resolution of approximately 2 feet per pixel and containing segments of a multilane freeway, with a few vehicles and road surface markings (painted arrows and words in the extreme left lane). The variations in road surface materials, centerlines, and intralane wear patterns correspond linearly to the road itself. The vehicles and other anomalies, however, stand out as being quite different from the pattern of the road.

These observations lead to a simple model for the photometry of most of the road surface. The model predicts that successive road reflectance profiles extracted perpendicularly to the direction of the road and centered upon it should have approximately the same appearance. Deviations from the model, defined as anomalies, are caused by road



FIGURE 14 A FREEWAY INTERSECTION TEST SITE

surface markings, vehicles, shadows, changes in road width and constituent materials, and by other influencing factors. The first attempt at a road-tracking algorithm exploited this model. Successive road reflectance profiles (RRP) extracted perpendicular to the direction of the road showed a high degree of correlation, which suggested that road tracking could be accomplished by using a cross-correlation-based approach. The location of the correlation peak was used to maintain alignment with the road center and to generate a model for the road trajectory. However, this first attempt turned out to be unsatisfactory because anomalies perturbed the location of the correlation peak and, concomitantly, other small errors in locating the correlation peak accumulated.

To overcome these problems, six refinements were introduced:

* A cumulative road reflectance profile model
* Bounded-difference alignment
* Masked alignment
* Match-function peak interpolation
* Anomaly detection
* Trajectory extrapolation.

The cumulative road reflectance profile model was introduced to reduce the tendency of alignment errors between successive RRPs to cause an increasing drift away from the road center. Each new RRP is aligned with the current RRP model, rather than with the preceding RRP. The RRP model is an exponentially weighted average of all RRPs previously encountered in a road segment (excluding anomalous points), as expressed in the following equation:

$$RRPmodel(t+1,i) = K*RRPmodel(t,i) + (1-K)*RRP(t+1,i)$$

K determines the "half-life" of the model. The choice of exponential weighted average rather than some other form of average was empirical. The model is initialized to the first RRP.

Bounded-difference alignment is a technique which is less prone than conventional cross correlation alignment to cause misalignment of RRPs. This is accomplished by computing the mean bounded absolute difference between two RRPs for a variety of alignments according to the following formula:

```
find k which minimizes summation over i of
    min( KMAD, abs[ RRPmodel(t,i)-RRP(t,i+k) ] )
```

where KMAD is set to a multiple (like 1.5 or 2) of the expected mean absolute difference between RRPs which do not contain anomalies.

Masked alignment further reduces the effect of anomalies on the alignment peak by eliminating from the alignment function those points whose absolute difference exceeds a threshold similar to KMAD. Match-function peak interpolation uses parabolic interpolation to refine to sub-pixel accuracy the estimate of the alignment peak location.

Anomaly detection is accomplished by comparing the aligned RRP with the RRP model. Corresponding pixels that disagree by more than a threshold (similar to KMAD) are marked as anomalies. Anomaly detection can be done densely, so that every pixel on the road is tested against the RRP model. This differs from RRP alignment, which is performed only as needed to determine the course of the road.

Parabolic extrapolation of the locations of previous road centers is used to predict road trajectory. The trajectory prognosis is used to guide the tracker past areas where the match function peak is unsatisfactory or an inordinate portion of the road consists of anomalies.

$$x(l) = a*l^2 + b*l + c$$
$$y(l) = d*l^2 + e*l + f$$

where l is length of road path, and a-f are coefficients determined by a least-squares approximation of the road course over the few (typically six) preceding RRP alignment points.

Steps for the refined tracking algorithm are given below:

(1) Based on past road center points and directions, extrapolate the position of the road center K feet ahead, using the parabolic model.

(2) Extract the road reflectance profiles (RRP) along a line perpendicular to the direction of the road at the extrapolated center point.

(3) Use mean bounded absolute-difference alignment to determine displacement of the current RRP with respect to the RRP model.

(4) Generate a mask indicating the positions of anomalous pixels that deviate from the RRP model.

(5) Use masked alignment to locate more accurately the proper alignment of the RRP with the RRP model.

(6) Use match-function peak interpolation to refine the alignment.

(7) Detect the anomalies by comparing each pixel in the vicinity of the current RRP with the RRP model.

(8) Update the RRP model, using only the "good" points of the current RRP at the optimum alignment. Updating is done using the exponentially weighted average.

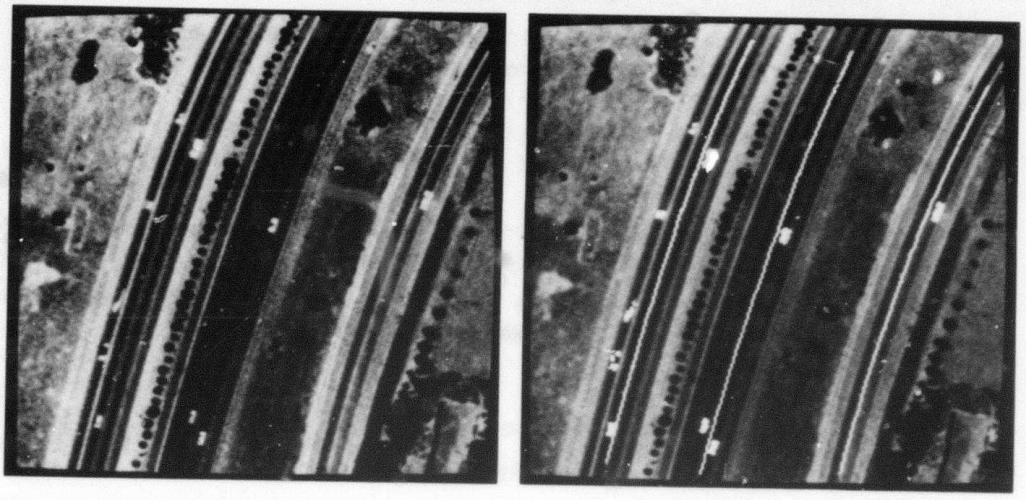
(9) Repeat steps 1-8 until the edge of the image is encountered or the RRP model becomes invalid (see the following section).

B. Experimental Results

In the experiments shown here, the road tracker was interactively started by indicating the following information for each road segment:

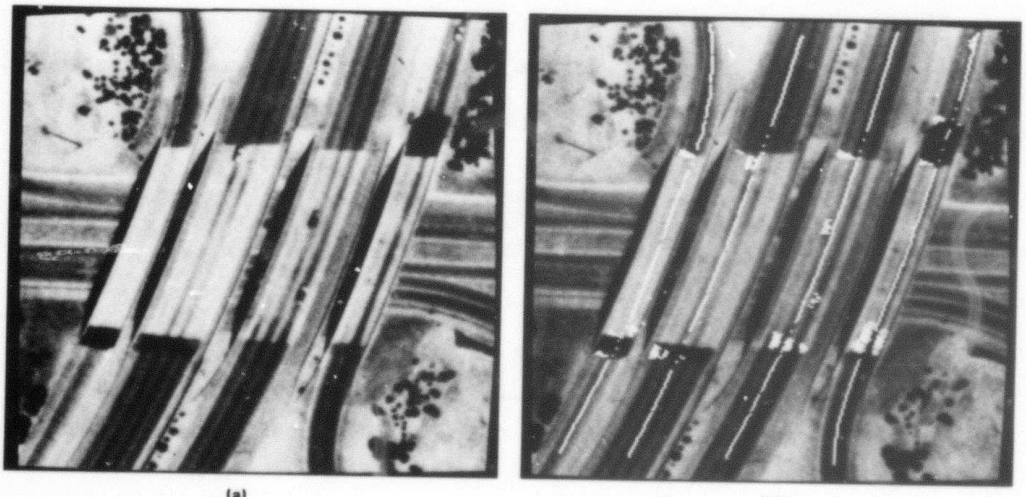
<X0,Y0>	center of road lane
theta0	direction of road at <X0,Y0>
w	nominal width of road

The freeway example in Figure 15 conforms well to the above road model, as shown by the overlay results in Figure 15b. The bright lines indicate the road trajectory and the bright blobs indicate potential anomalies.



(a) (b)

FIGURE 15 A SECTION OF FREEWAY



(a) (b)

FIGURE 16 A FREEWAY OVERPASS

The simplistic model in which a road consists of well-correlated reflectance profiles clearly breaks down in the example shown in Figure 16a, where the road surface changes from concrete to asphalt on the overpass. Certainly the RRP model generated for the asphalt will not match the intensities in this different road surface.

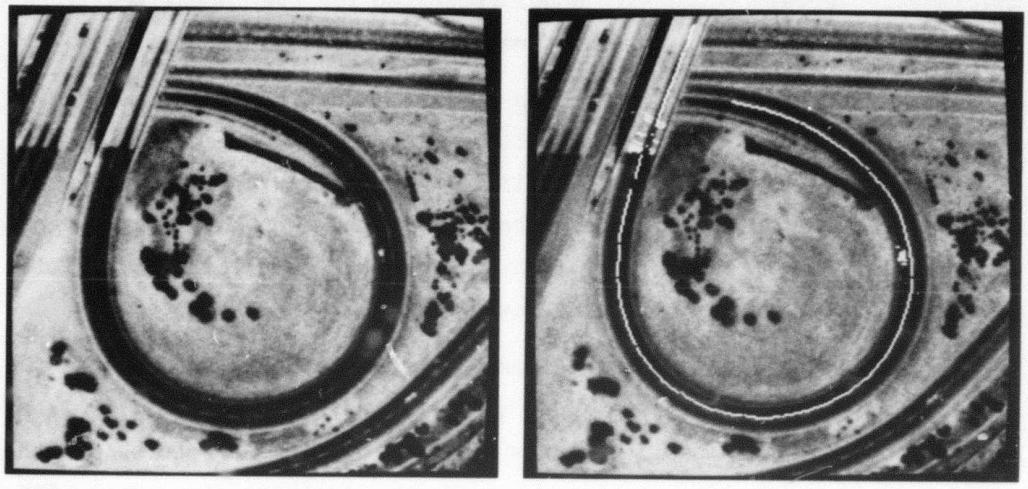
When the tracker encounters the surface change a high percentage of the pixels in the RRP will be anomalous (Figure 16b). When this occurs, the tracker extrapolates in anticipation and tries to reacquire the road. If the road is not reacquired within the length of the longest expected anomaly, the tracker then assumes that a pavement transition has taken place and establishes a new RRP model.

The four vehicles shown in Figure 16 were detected, but most of the anomalies marked therein are due to road surface changes. A later section will discuss basic changes in the control structure of the current program to eliminate the false alarms resulting from the surface changes.

Figure 17 shows results for a freeway interchange on-ramp loop. This example is interesting because the road curves rather tightly, and the road surface changes at approximately the same place where the road trajectory changes from a circular arc to a straight line.

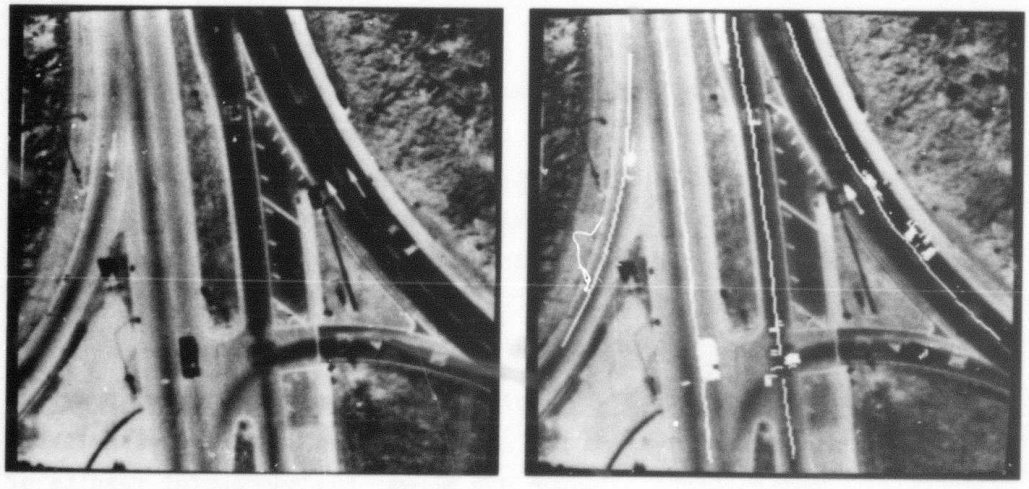
Figure 18 illustrates a very complicated example of road forks, variation in lane width, and intersections. For the lanes tracked all vehicles and at least portions of the road surface marks (arrows and words) were found. In a developed road-expert system, the data base should help significantly in handling the complexities of this image through its knowledge of the location of forks, intersections, lane-width changes, and the like. This information will facilitate the task of interpreting causal factors in RRP model changes.

In marked contrast to the situation in most of the previous figures, Figure 19a shows a rather poorly defined dirt road with little evidence of wear patterns. Figure 19b shows the successful results of the road tracker. Most of the anomalies marked were due to shadows cast by sparsely foliaged trees.



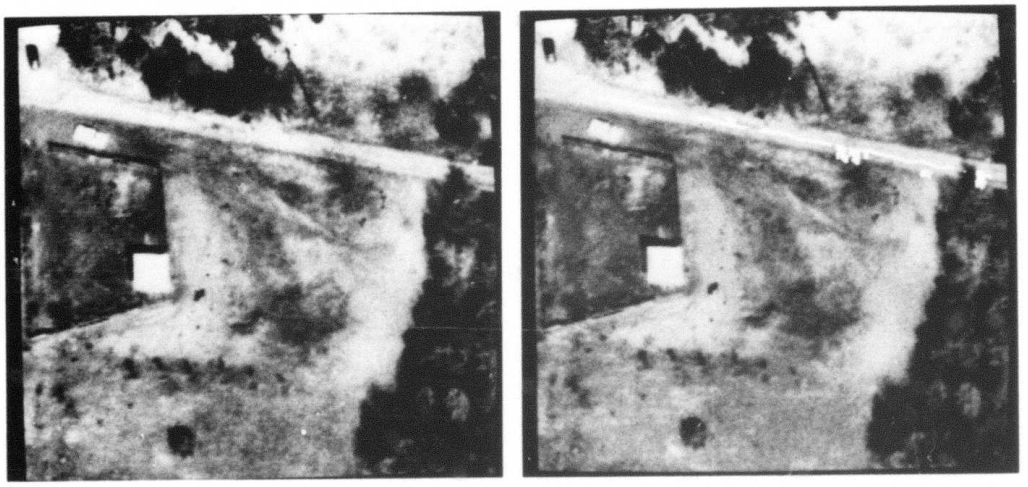
(a) (b)

FIGURE 17 AN INTERCHANGE ON-RAMP



(a) (b)

FIGURE 18 A BUSY INTERSECTION



(a) (b)

FIGURE 19 A DIRT ROAD

C. Concluding Discussion

The preceding examples demonstrate both the capabilities and limitations of the present tracking algorithm. The algorithm has shown surprising ability to cope with a wide variety of road situations, including total change in the road surface. The use of bounded-difference and masked-alignment techniques nearly eliminates perturbation of the road track by anomalies. Trajectory extrapolation enables the tracker to reacquire the road after detecting that the road surface has changed. All results were obtained using the identical program and the same detection and threshold criteria; no attempt was made to "fine-tune" the parameters individually for each example.

One defect of the present algorithm is the attempt to accomplish too much in one pass along the road. In particular, anomaly marking, in the present system, begins before road-surface changes have been detected. The false alarms induced by this defect can be eliminated either by backtracking when a road transition is found, or by performing the detailed anomaly detection in a second pass along the road, utilizing the road-course and surface-change knowledge generated by the tracker.

The road tracker presently operates as an independent module. As a component of a larger road-expert system, it will be initiated from the output of a map-guided road-detection algorithm operating on lower-resolution imagery. Data-base knowledge can also be used in the tracking algorithm to increase reliability and reduce false alarms in anomaly detection. Such knowledge might consist of previous imagery of the same area or geometric information about the location of road forks, intersections, overpasses, surface changes, lane-width changes, and other parameters. To exploit such knowledge, it is necessary to establish geometric correspondence between the image and the data-base coordinate system. If, for example, a road anomaly corresponds to a known surface marking on the map or appeared in the same place in previous images, it is probably a surface marking rather than a vehicle. Similarly, the use of an illumination model can help to distinguish shadow-casting objects from surface markings.

We have acquired and digitized images taken under diverse viewing conditions with seasonal variations. We have also developed algorithms to introduce clouds and cloud shadows into these images to simulate realistic situations in which visibility is impaired and surface lighting modified. This will permit testing of the tracking algorithms under controlled conditions of nonvisibility of road segments (due to clouds) and major photometric differences between key features in images of the same area. The use of a map data base will be essential to guide the interpretation of such images; as noted earlier, however, determining how such apriori knowledge should be used is a major focus of our research effort.

With the enhancements and improvements that are planned, it should be possible to track roads and detect potential vehicles with very high hit rates and low false-alarm rates, even when operating with difficult imagery. This capability is a central component of an overall road-monitoring system.

Annex E

Knowledge-Based Detection and Classification of Vehicles and other Objects in Aerial Road Images

(From "Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, SRI Project 5300, May 1979, pp. 10-25)

II KNOWLEDGE-BASED DETECTION AND CLASSIFICATION OF VEHICLES AND OTHER OBJECTS IN AERIAL ROAD IMAGES

A. Introduction

This section describes an approach to finding and identifying vehicles in aerial images, using diverse sources of knowledge. The following scenario provides a context for this work. Given a digital aerial image and a data base, the problem is to detect vehicles on the road and to classify them as to vehicle type. The image should have sufficient spatial resolution to allow recognition (about one ft. per pixel, minimum). Figure 1 shows a typical image of an area containing a freeway interchange.

The data base contains information about some limited geographical area of interest. As a minimum, it should have the locations of known roads in the area. Other relevant information could include (but not be limited to):

* Road width
* Brightness profiles across the road
* Terrain information
* Buildings, railroads, and other cultural features
* Intersections, overpasses, and access roads
* Signs and permanent road markings
* Previous photo coverage of the area, in digital form.

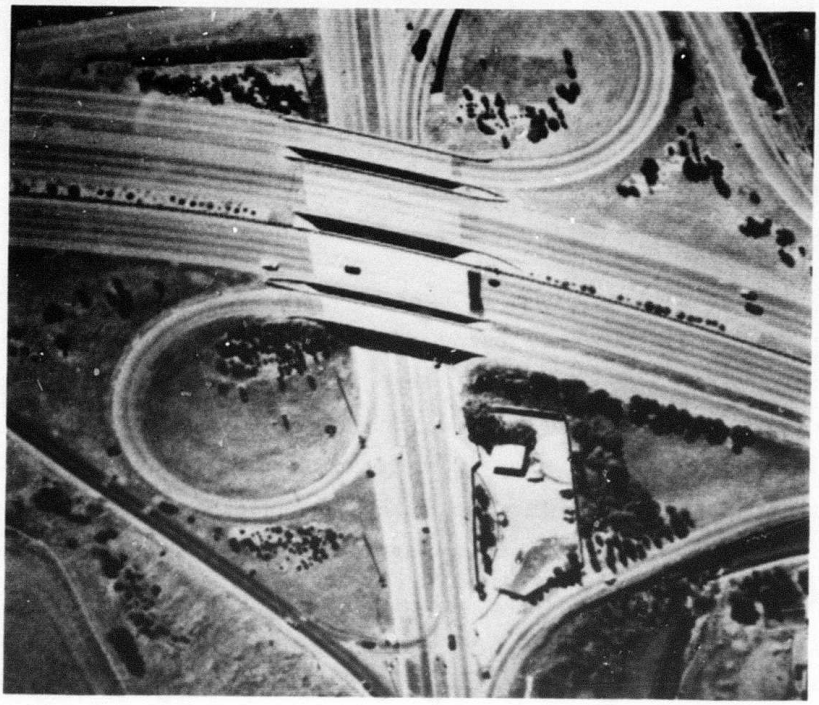


Figure 1 An Aerial Road Image

A calibration procedure [6] establishes correspondence between image coordinates and geographic coordinates, allowing us to convert quickly back and forth between coordinates in the data base and pixel locations in the image. A road tracker [3] uses the road location predicted by the data base to trace the road centerline and boundaries by correlating successive profiles perpendicular to the road direction. Areas where the image diverges from the expected road profile are identified as "anomalies." These areas are passed to the classification routines for further scrutiny.

Many different conditions could give rise to an anomaly. Vehicles usually show up this way, but so do the shadows of objects off the road (trees, buildings, signs, utility poles), overhanging trees, painted markings on the road, and changes or irregularities in the road surface

(such as tar patches). There are also some less frequent situations with which a practical system ought to deal, such as road construction, floods, bomb craters, smoke, and dust clouds. The classifier must first decide if the anomaly arises from a vehicle or from some other cause. Then it can classify the vehicle type.

Although the scenario assumes some rather specific resources and goals, this knowledge-based approach is generally applicable to a wide range of object recognition tasks in cartography and photo interpretation.

B. Sources of Information

A wide variety of information can be helpful for detecting and classifying vehicles. We can identify three kinds of knowledge relevant to this problem: about the problem domain (generic knowledge), about the site (the data base), and about a particular place and time (information associated with the image).

Generic knowledge includes information that can be deduced from functional descriptions. A road is a narrow, linear region upon which vehicles may travel. The road is usually continuous in the image--if it appears discontinuous it may be that there are obstructions, or there may be shadows or discolorations on the road surface. Roads have minimal variation in the direction of travel but may have considerable variation in the perpendicular direction, because of the different compositions of roadbed, shoulders, and an expected pattern of oil stains in the center of each lane. We have some idea of the expected shapes of vehicles viewed from different angles, and an expectation that they probably will be aligned parallel to the road direction. Our illumination models take into account the physics and geometry of shadows, and we can sometimes use shadows to draw inferences about objects. We know the usual places where road signs, utility poles, and painted road markings are located. All the foregoing can be used to make sense out of a road scene.

The data base is a useful source of information. Its principal use is to predict the approximate road centerline, so that the road-tracking subroutines can operate. But other kinds of information can be brought into play. Terrain information can be used to refine position estimates when the viewing angle is not vertical and to predict shadows better if the ground slopes. Classifying shadows of objects off the road is very much simplified when it is known what objects are likely to cast shadows. Ambiguous anomalies in the image can sometimes be distinguished if a picture can be compared with a previous one or, better yet, if the data base states what anomalies were found in previous images and how they were classified. Intelligence reports and expected traffic conditions can help the program decide what to look for or what strategies to use.

The greatest single source of data is the image itself. It is easy to overlook some information that is associated with the image but may not be in the actual raster. For example, it is usually possible to ascertain (at least approximately) the altitude, position, and heading of the aircraft from which the image was taken. Scaling parameters, view angles, and compass headings can be derived by calibration. If the time and date of the picture are known, the sun position can be calculated--but even without these data the sun position usually can be estimated from shadows.

In short, detection and classification of vehicles are not based solely on what is in the image. In the following sections, we detail some of the ways we use the available information.

C. Use of the Correlation Road Tracker

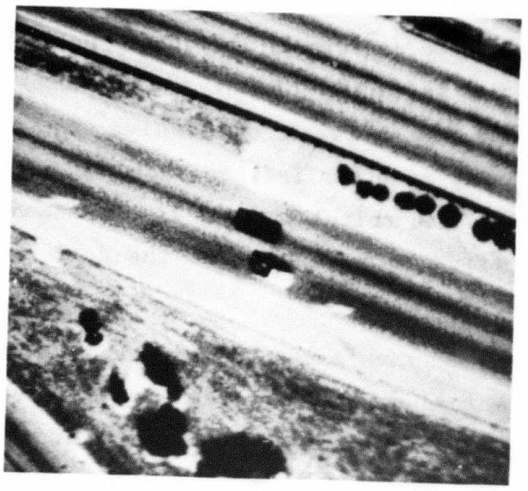
We depend on the correlation road tracker designed by Quam [3] to isolate anomalies in images of roads. These are regions where attention should be focused.

The road tracker is based on the assumption that variations in road surface materials, centerlines, and intralane wear patterns correspond linearly to the road itself. Vehicles and other anomalies, however,

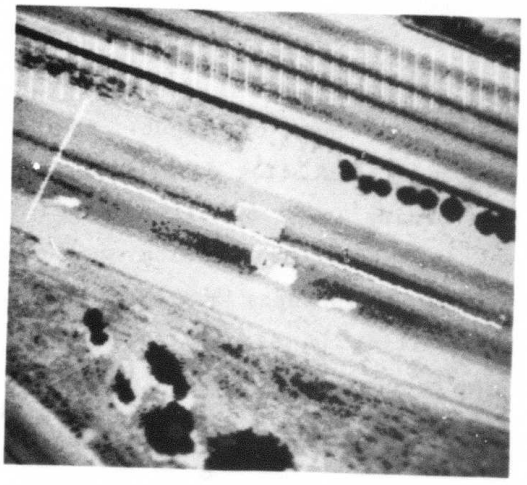
stand out in sharp contrast to the pattern of the road. Detecting these anomalies is important to the operation of the road tracker. Where substantial disagreement occurs between successive profiles, the corresponding pixels are marked as anomalies, so that these points can be eliminated from the correlation calculations. If the anomalies were not so masked, they would perturb the location of the correlation peak and introduce errors.

Figure 2a shows a representative excerpt from the area covered by the image of Figure 1. The road tracker is initiated by specifying a single profile approximately perpendicular to the road direction and centered on it. This initial baseline is now selected manually, but facilities exist for using the data base to draw the baseline automatically.

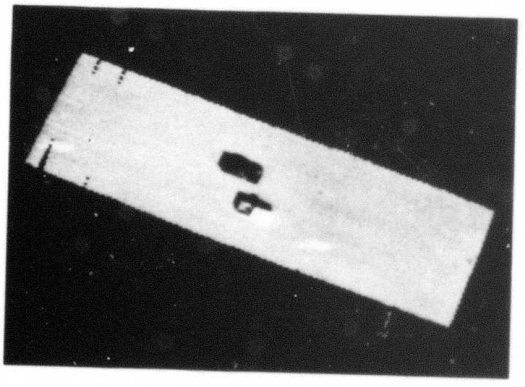
The road tracker produces several forms of output. As indicated by Quam [3], the program can produce a point list describing the track of the road center, as well as a binary image of all points in the road that are anomalous. But for vehicle identification another form of output has been added. The road reflectance model may be subtracted from each pixel considered, resulting in a difference image that has been normalized to remove the road profile. Figure 2b shows the baseline, the road center, and anomalies detected. Figure 2c shows the difference image. The difference image may be converted to a binary anomaly image by thresholding.



(a) ROAD SCENE



(b) BASELINE, CENTERLINE, AND ANOMALIES



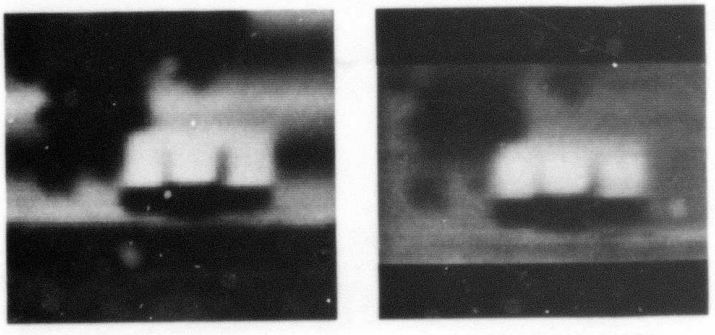
(c) DIFFERENCE IMAGE

Figure 2 Operation of the Road Tracker

In the difference image, shadows tend to have a relatively uniform intensity, even though the road reflectance profile varies considerably. If we adopt the simplifying assumptions that any object casting a shadow may be approximated by a half plane of infinite extent that hides all but a fixed proportion of the sky, and if we neglect reflected illumination from nearby objects, then the ratio of intensities across the shadow edge should not depend on the reflectivity of the underlying surface. When the original image is digitized on a logarithmic brightness scale, this constant ratio becomes a constant intensity in the difference image. Because the assumptions are approximate at best, the constant-difference test is almost never exact. Nonetheless, by subtracting the road profile from the image, we can expect the intensity of shadows to be more uniform in the difference image than in the original one.

On the other hand, when anomalies are caused by vehicles, subtracting the road profile will cause its inverse to be superimposed on the anomaly. Figures 3a and b show an original image and a difference image (from another road site) that demonstrate these peculiarities. Both kinds of image are useful in classifying anomalies.

As the road tracker proceeds, it constantly keeps track of the average correlation between successive road profiles at their optimum locations. This correlation value, a useful estimate of noise in the picture, is made available to succeeding classification stages.



(a) ORIGINAL IMAGE (b) DIFFERENCE IMAGE

Figure 3 Original and Difference Image

D. Shadows

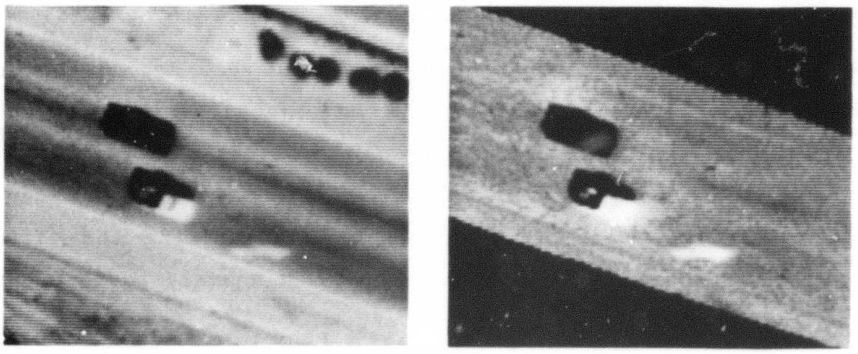
An understanding of shadows is crucial to making sense out of high-resolution aerial images. The scene is always out-of-doors and usually illuminated by direct sunlight, which produces deep, dark shadows. Frequently shadows are the most prominent visual feature of an image.

For vehicle classification, many of the anomalies the classifier is called on to consider are the shadows of objects off the road, such as trees, signs, or utility poles. All vehicles cast shadows, and, unless the boundary between the vehicle and its shadow can be determined, classification on the basis of shape is hopeless. Furthermore the existence or nonexistence of a shadow can aid in deciding whether or not a given anomaly is a vehicle. The size and shape of the shadow can give valuable clues as to the height of the vehicle and its profile. As a dramatic demonstration of this, consider the vehicle shown in Figure 4. Because its reflectance is almost the same as that of the road, the vehicle might have gone unnoticed, were it not for the shadow. But the shadow not only gives away its position; it tells us the vehicle is probably a Volkswagen "beetle."



Figure 4 Vehicle with Shadow

We have a number of techniques at our disposal for identifying shadows. The simplest is based on the brightness model. The technique is simply to search for all pixels in the image whose intensity is in the range of values expected for shadows. This works somewhat better in the difference image than in the original, because the effects of variation in the road surface are reduced. Figure 5 shows the central portion of the area analyzed in Figure 3, which we shall use to illustrate shadow-finding techniques. Figure 6 shows the shadows extracted from Figure 5b by this method.



(a) ORIGINAL IMAGE (b) DIFFERENCE IMAGE

Figure 5 Original and Difference Pictures

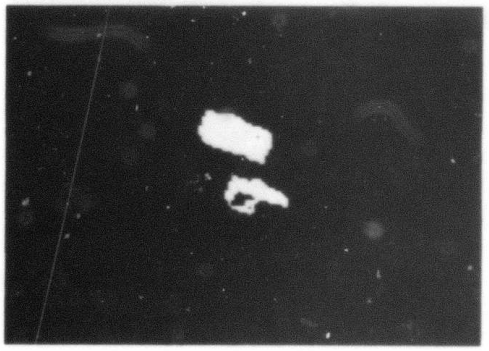


Figure 6 Shadows Found by Brightness Criterion

In our work so far, the expected range of shadow intensities has been inferred from the statistics of areas manually indicated as shadows. It should be possible in principle to automate this procedure--for example, by using the using the data base to predict or find known shadows. Alternatively, it seems likely that a formula can be derived that will give the expected distribution based on calibration of photometry.

In situations in which the correlation road tracker is not applicable, shadows located by the brightness model might indicate areas of the picture that merit scrutiny.

Another device, based upon a predictive model, depends on knowing the sun's angle. The shadow of any raised object is always on the side away from the sun; and, if the height of the object is known, the length of the shadow can be predicted. Figure 7 shows the areas identified as shadow from the image of Figure 5b by thresholding the difference image to locate anomalies and by assuming each anomaly to be due solely to an object five ft. tall, plus its shadow.

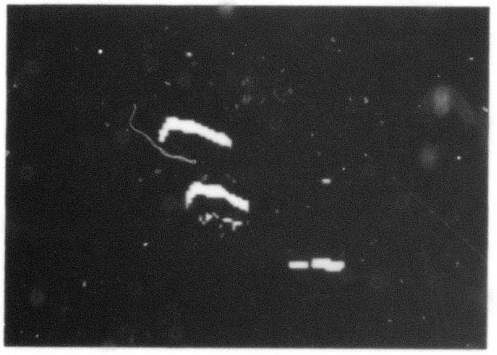


Figure 7 Shadows Found by Predictive Criterion

The third technique is based on a projective model. It tries to look directly for the shadow edge. Vehicles tend to be rectangular when viewed from above; and, unless the sun is directly ahead of or behind the vehicle, there will be a long, straight edge separating the vehicle

from its shadow. This edge can usually be found by performing a Hough transform [7] on the gradient of the image, or, equivalently, by projecting the gradient onto axes oriented in various directions and finding the direction from which the gradient points tend most to reinforce one another. However, much better results are obtainable when the direction of the edge is known or assumed a priori. Such is usually the case, for vehicles tend to be oriented parallel to the road direction.

An example of shadow detection by projection is presented in the next section.

The three techniques are based on different sets of assumptions and are applicable in different circumstances. The projective method is useful only for finding shadows of vehicles. The predictive model is more generally useful, being applicable to objects off the road as well as on it. The brightness model makes no assumptions about the object casting the shadow--it only requires that the background on which the shadow is cast be relatively uniform.

E. Classification of Anomalies

For classifying anomalies, we have chosen to construct a number of "expert" subroutines, each of which tests a specific hypothesis. For example, the vehicle expert determines whether or not a given anomaly could be a vehicle (plus its shadow) and if so, attempts to distinguish whether the vehicle is a car or a truck. The tree shadow expert tries to say whether or not the anomaly could be the shadow of an object off the road, and the road marking expert similarly looks for painted markings. Other expert modules could easily be integrated into the scheme. The experts operate in parallel, each expert forming its decision without interacting with its counterparts. The top-level program chooses the most likely interpretation of the anomaly. If no expert subroutine is able to account for the anomaly, it is labeled "unclassified."

The vehicle expert is the most involved of the expert subroutines. It first examines the overall size (area) of an anomaly. If the anomaly is too small or too large, it is rejected. Next, by projecting the gradient image to a baseline, long edges are found that might correspond to sides of the car. A binary mask is used for the projection, so that only those points near the anomaly are considered; the mask is generated by expanding ("growing") the anomaly region by three pixels. Figure 8a shows the results of applying a gradient operator to the image of Figure 5a. The masked gradient was projected on the axis drawn in Figure 8b, where the average projected gradient magnitude is plotted.

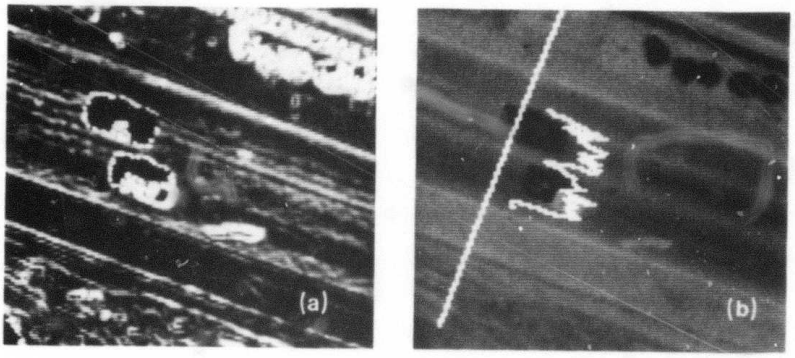


Figure 8 Use of Projection to Find Shadow Edges

A line perpendicular to the direction of the road is used as an initial baseline. If some evidence of edges is found, the orientation is perturbed a small amount to find a local maximum. If the edges are not found, a global search is made for a direction of projection that will show the edges. If the edges are again not found, the anomaly is rejected.

Note that there are three peaks in the plot, corresponding to the boundaries between road and car, between car and shadow, and between shadow and road. The three highest peaks in the projected gradient are examined to see if they are in the correct relationship. Average brightness is projected to the same baseline to see if the brightness of the shadow portion is appropriate. A figure of merit is computed from

these tests, indicating the degree to which the measured spacing and brightness approximate the expected spacing and brightness. The figure of merit is used later in choosing the most likely interpretation of the anomaly.

The average length of the shadow and the location of the sun may be used to estimate the height of the vehicle. A tolerance or range of uncertainty is also computed at this time, because the combination of low spatial resolution and a disadvantageous sun angle may make the height figure not particularly useful. A nominal height of 6 ft. is used for predicting a shadow to the front or the rear of the vehicle; this predicted shadow length subtracted from the length of the original anomaly yields the length of the vehicle.

Classification as to vehicle type is relatively crude at this time. If the overall length of the vehicle is greater than 20 ft., or if the height can be reliably stated as exceeding 6 ft., the vehicle is called a "truck." Otherwise it is called a "car."

Another expert subroutine identifies shadows of objects off the road. To qualify as such a shadow, an anomaly must have an average brightness lower than the average road brightness and extend to the edge of the road on the side nearer the sun. A figure of merit is calculated from the extent to which the average brightness (in the difference image) corresponds to the predicted value, as well as from the variance of brightness inside the anomaly.

The expert on painted road markings is similar to the shadow expert. Painted markings are always brighter than the road surface and limited in total area. The figure of merit is based only on variance of brightness; a much lower variance is expected for road markings than for shadows.

F. Discussion

The state of our experiments in anomaly classification is such that it is too early to report any quantitative results. However, we can say, qualitatively at least, that the methods outlined above succeed in the easy cases and break down for the difficult ones. We have tested our programs on approximately 20 different scenes extracted from three diverse road areas. Where good contrast exists between an anomaly and the road, and (in the case of vehicles) the shadow is visually distinct from the object casting it, we have little difficulty in obtaining a correct identification. Where conditions are not as good, the programs tend to make no identification at all, rather than come up with a misclassification. Additional robustness in the classifier will be necessary to enable it to handle unusual cases.

The various expert subroutines are not now integrated in any way. Each reports its figure of merit to the top-level program, which selects among the hypotheses. A more useful system should allow interaction among the various experts.

Figure 2 shows a good example of a case that could be handled by cooperation of the tree-shadow and the vehicle experts. It might be sufficient if the shadow expert were to realize that it could interpret part of the anomaly, subtract the explainable part, and ask the other experts to classify what remains. The vehicle expert would have to take the situation into account and not look for a separate shadow for this anomaly.

Figure 9 is difficult to analyze without higher-level knowledge. A more direct link to the data base would be particularly useful in this case, enabling us to divide the anomaly into portions that are "expected" (the visible portions of the arrow) and "not expected" (the car and its shadow).

Much generic knowledge tends to be expressed in the coding of the computer programs that analyze pictures. In this form it is inflexible--adding new knowledge involves writing new computer programs. A long-

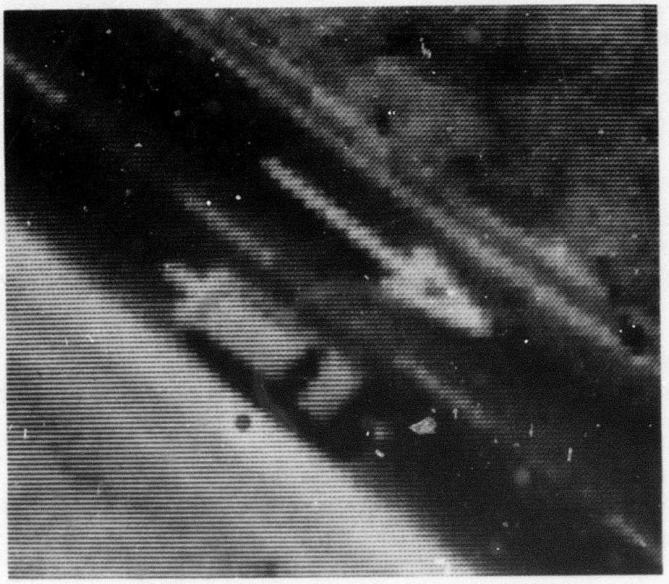


Figure 9 A Vehicle over a Road Marking

range goal of this research is to find new ways of expressing this kind of information--for example, in the form of rules or templates. Such a capability would lead to highly competent computer visual capabilities that would greatly enhance interactive and automatic cartography and photo interpretation.

REFERENCES

1. H. G. Barrow et al., "Interactive Aids for Cartography and Photo Interpretation: Progress Report, October 1977," Proceedings: Image Understanding Workshop, pp. 111-127 (October 1977).

2. M. A.Fischler et al., "Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, SRI Project 5300, SRI International, Menlo Park, California (October 1978).

3. L. H. Quam, "Road Tracking and Anomaly Detection," Proceedings: ARPA Image Understanding Workshop, pp. 51-55 (May 1978).

4. G. J. Agin, "Knowledge-Based Detection and Classification of Vehicles and Other Objects in Aerial Road Images," Proceedings: ARPA Image Understanding Workshop (in press, April 1979).

5. M. A. Fischler et al., "The SRI Image Understanding Program," Proceedings: ARPA Image Understanding Workshop, Menlo Park, California (April 1979).

6. R. C. Bolles et al., "The SRI Road Expert: Image-to-Data-Base Correspondence," Proceedings: ARPA Image Understanding Workshop, Pittsburgh, Pennsylvania (November 1978).

7. R. O. Duda and P. E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures," Communications of the ACM, Vol. 15, No. 1 (January 1972).

Annex F

Reports and Publications

Annex F

Reports and Publications

Publications Resulting from DARPA IU Contracts

Conference Papers:

[C1] "Application of Interactive Scene Analysis Techniques to Cartography," T. D. Garvey and J. M. Tenenbaum, Proceedings of Third International Joint Conference on Pattern Recognition, Coronado, CA, November 1976.

[C2] "An Experiment with a System for Locating Objects in Multisensory Images," T. D. Garvey, Proceedings of Third International Joint Conference on Pattern Recognition, Coronado, CA, November 1976.

[C3] "IGS: A Paradigm for Integrating Image Segmentation and Interpretation," J. M. Tenenbaum and H. G. Barrow, Proceedings of Third International Joint Conference on Pattern Recognition, Coronado, CA, November 1976.

[C4] "A Geometrically-Indexed Knowledge Base for Understanding Aerial Imagery," H. G. Barrow, Proceedings of IEEE Workshop on Picture Data Description and Management, Chicago, ILL, April 1977.

[C5] "Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching," H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, Proceedings 5th IJCAI, MIT, Cambridge, MA, August 1977.

[C6] "Experiments in Map-Guided Photo Interpretation," H. G. Barrow, R. C. Bolles, T. D. Garvey, J. H. Kremers, J. M. Tenenbaum, and H. C. Wolf, Proceedings 5th IJCAI, MIT, Cambridge MA, August 1977.

[C7] "Automatic Determination of Image-to-Data-Base Correspondence," R. C. Bolles, L. H. Quam, M. A. Fischler, and H. C. Wolf, Proceedings 6th IJCAI, Tokyo, Japan, August 1979.

[C8] "Detection of Roads and Linear Structures in Aerial Imagery by Computer," M. A. Fischler, J. M. Tenenbaum, and H. C. Wolf, SRI Technical Note 200, October 1979; to be published in Computer Graphics and Image Processing.

[C9] "A Computer-Based Approach to Perceptual Reasoning and Multisensor Integration," M. A. Fischler, T. D. Garvey (in preparation).

[C10] "Map-Guided Interpretation of Remotely-Sensed Imagery," J. M. Tenenbaum, H. G. Barrow, R. C. Bolles, M. A. Fischler, and H. C. Wolf, Proceedings of IEEE Conference on Pattern Recognition and Image Processing, Chicago, ILL, August 1979 (in press).

Reports Submitted to DARPA

Semiannual Reports:

June 1976 [R1]
"Interactive Aids for Cartography and Photo Interpretation," H. G. Barrow, T. D. Garvey, J. H. Kremers, J. M. Tenenbaum, and H. C. Wolf, in Artificial Intelligence--Research and Applications Annual Progress Report, ARPA Order No. 2894, Contracts DAHC04-75-C-0005 and DAAG29-76-C-0012 Ed B. Raphael

Nov 1976 [R2]
"Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, ARPA Order No. 2894-5 Contract DAAG29-76-C-0057 H. G. Barrow

May 1977 [R3]
"Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, ARPA Order No. 2894-5 Contract DAAG29-76-C-0057 H. G. Barrow

Oct 1977 [R4]
"Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, ARPA Order No. 2894-5 Contract DAAG29-76-C-0057 H. G. Barrow

June 1978 [R5]
"Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, ARPA Order No. 2894-5 Contract DAAG29-76-C-0057 G. J. Agin, H. G. Barrow, R. C. Bolles, M. A. Fischler, T. D. Garvey, J. H. Kremers, L. H. Quam, J. M. Tenenbaum, and H. C. Wolf

Oct 1978 [R6]
"Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, ARPA Order No. 2894-5 Contract DAAG29-76-C-0057 M. A. Fischler, G. J. Agin, H. G. Barrow, R. C. Bolles, L. H. Quam, J. M. Tenenbaum, and H. C. Wolf

May 1979 [R7]
"Interactive Aids for Cartography and Photo Interpretation," Semiannual Technical Report, ARPA Order No. 2894-5 Contract DAAG29-76-C-0057

M. A. Fischler, G. J. Agin, H. G. Barrow, R. C. Bolles, L. H. Quam, J. M. Tenenbaum, and H. C. Wolf.

IU Workshop Papers

6-7 Mar 1975 [IU1] "Research on Image Understanding Systems"
R. O. Duda and J. M. Tenenbaum

12-13 Apr 1976 [IU2] "Stanford Research Institute Presentation"
H. G. Barrow and J. M. Tenenbaum

13-14 Oct 1976 [IU3] "Stanford Research Institute Presentation"
H. G. Barrow

20-21 Apr 1977 [IU4] "Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching"
H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, H. C. Wolf

[IU5] "Interactive Aids for Cartography and Photo Interpretation: Progress Report," April 1976 to April 1977
H. G. Barrow

20-21 Oct 1977 [IU6] "Interactive Aids for Cartography and Photo Interpretation: Progress Report," October 1977
H. G. Barrow

3-4 May 1978 [IU7] "Road Tracking and Anomaly Detection in Aerial Imagery"
L. H. Quam

[IU8] "An Expert System for Detecting and Interpreting Road Events Depicted in Aerial Imagery"
H. G. Barrow and M. A. Fischler

14-15 Nov 1978 [IU9] "The SRI Road Expert: An Overview"
M. A. Fischler, G. J. Agin, H. G. Barrow, R. C. Bolles, L. H. Quam, J. M. Tenenbaum, and H. C. Wolf

[IU10] "The SRI Road Expert: Image-to-Data-Base Correspondence"
R. C. Bolles, L. H. Quam, M. A. Fischler, and H. C. Wolf

24-25 Apr 1979 [IU11] "The SRI Image Understanding Program,"
M. A. Fischler, G. J. Agin, H. G. Barrow, R. C. Bolles, L. H. Quam, J. M. Tenenbaum, and H. C. Wolf

[IU12] "Knowledge-Based Detection and Classification of Vehicles and Other Objects in Aerial Road Images"
G. J. Agin

SRI Technical Notes

Jan 1977 [S1] 137: "Interactive Aids for Cartography and Photo Interpretation,"
H. G. Barrow, T. D. Garvey, J. H. Kremers, J. M. Tenenbaum, and H. C. Wolf

April 1977 [S2] 153: "Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching,"
H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, H. C. Wolf

Other Publications

[P1] "Representation and Use of Knowledge in Vision"
H. G. Barrow and J. M. Tenenbaum,
in Naval Research Reviews, Vol. 29, No. 4, pp. 3-17 (1976).

[P2] "Experiments with Interpretation-Guided Segmentation"
J. M. Tenenbaum and H. G. Barrow,
Artificial Intelligence, Vol. 8, No. 3, (North Holland, June 1977).

[P3] "IGS: A Paradigm for Integrating Image Segmentation and Interpretation"
J. M. Tenenbaum and H. G. Barrow,
in Pattern Recognition and Artificial Intelligence,
C. H. Chen, ed. (Academic Press, New York, New York, 1977).

Films

[F1] HAWKEYE
Director R. C. Bolles

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

INTERACTIVE AIDS FOR CARTOGRAPHY AND PHOTO INTERPRETATION

5. TYPE OF REPORT & PERIOD COVERED

Final Technical Report
5-12-76 to 10-9-79

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

M. A. Fischler, R. C. Bolles, H. C. Wolf, G. J. Agin, L. H. Quam, H. G. Barrow, J. M. Tenenbaum

8. CONTRACT OR GRANT NUMBER(s)

DAAG29-76-C-0057

9. PERFORMING ORGANIZATION NAME AND ADDRESS

SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709

12. REPORT DATE

December 1979

13. NO. OF PAGES

125

14. MONITORING AGENCY NAME & ADDRESS (if diff. from Controlling Office)

15. SECURITY CLASS. (of this report)

Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Image Analysis
Image Understanding
Automatic Photo Interpretation
Automated Cartography
Artificial Intelligence

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Research at SRI International under the ARPA Image Understanding Program was initiated to investigate ways in which diverse sources of knowledge might be brought to bear on the problem of analyzing and interpreting aerial images. The initial phase of research was exploratory and identified various means for exploiting knowledge in processing aerial photographs for such military applications as cartography, intelligence, weapon guidance, and targeting. A key concept is the use of a generalized digital map to guide the process

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19. KEY WORDS (Continued)

20 ABSTRACT (Continued)

of image analysis. The results of this earlier work were integrated in an interactive computer system called "Hawkeye." This system provides not only basic facilities necessary for a wide range of tasks in cartography and photo interpretation, but also a framework within which other applications can be readily demonstrated.

Since January 1978, work has been focused on development of a system (called the "SRI Road Expert") capable of expert performance in a specific task domain--road monitoring. This report summarizes the specific objectives, approach, and technical accomplishments of the research program.