

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.

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 SO₂ 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.



FIGURE 1 HIGH ALTITUDE VERTICAL MAPPING PHOTOGRAPH OF SAN FRANCISCO BAY AREA
Taken from a U-2 at 45,000 feet.

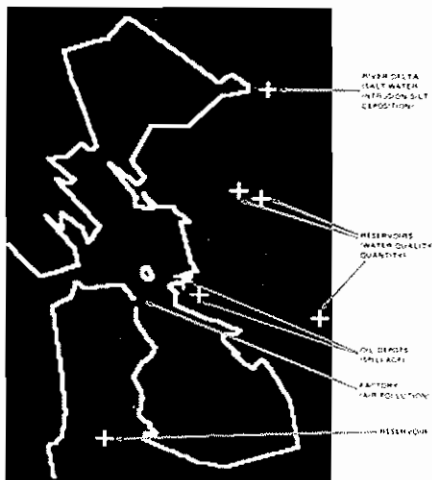


FIGURE 2 COMPUTER DISPLAY OF A SIMPLE MAP DATA BASE FOR THE SAN FRANCISCO BAY AREA SHOWING MAJOR LANDMARK (COASTLINE) AND REPRESENTATIVE MONITORING SITES (CROSSES)

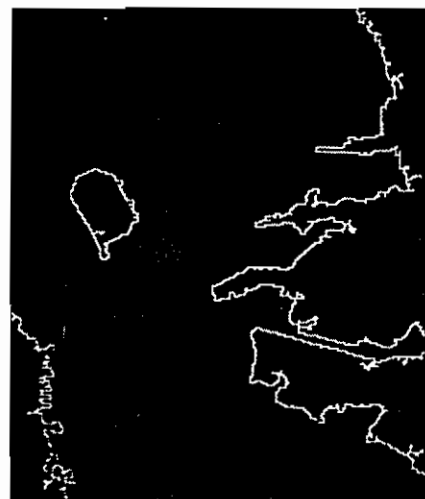


FIGURE 3 COASTLINE EXTRACTED BY BOUNDARY FOLLOWER



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

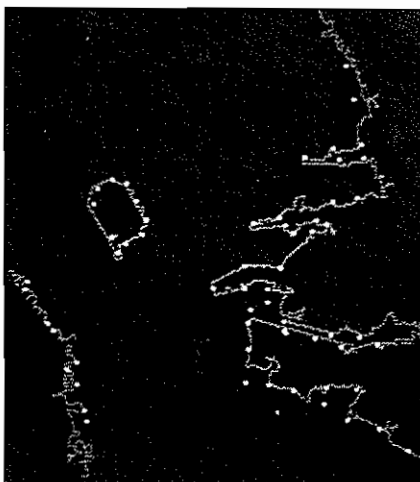


FIGURE 5 PREDICTED COASTAL COORDINATES AFTER OPTIMIZATION OF CAMERA PARAMETERS

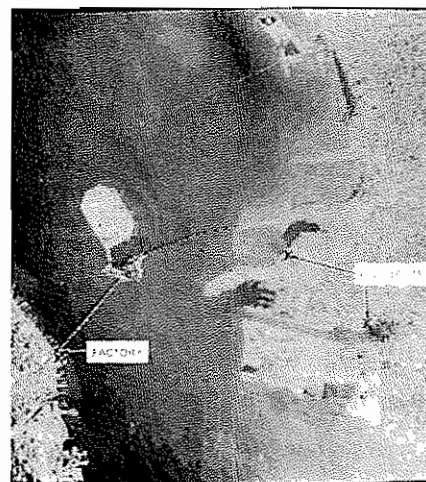


FIGURE 6 PREDICTED IMAGE LOCATIONS OF VISIBLE MONITORING SITES BASED ON OPTIMIZED PARAMETERS

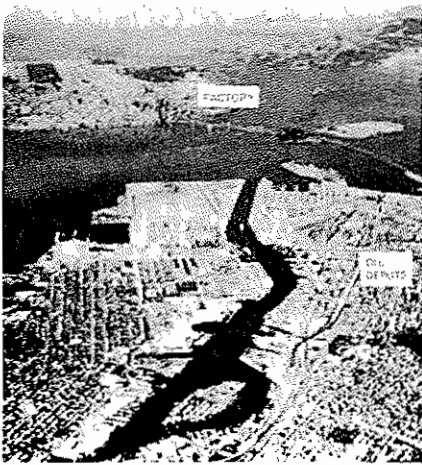


FIGURE 7 PREDICTED LOCATIONS OF VISIBLE MONITORING SITES IN AN OBLIQUE VIEW LOOKING WEST FROM ALAMEDA

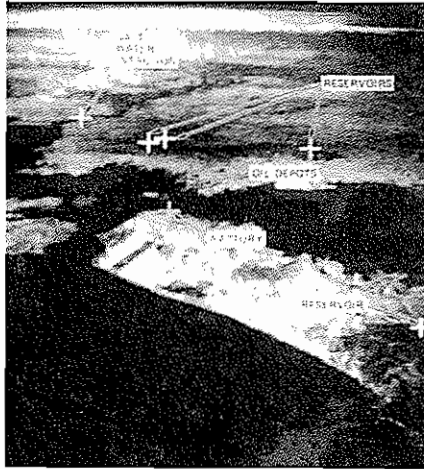


FIGURE 8 PREDICTED LOCATIONS OF VISIBLE MONITORING SITES IN A HIGH ALTITUDE OBLIQUE VIEW LOOKING EAST FROM THE PACIFIC OCEAN

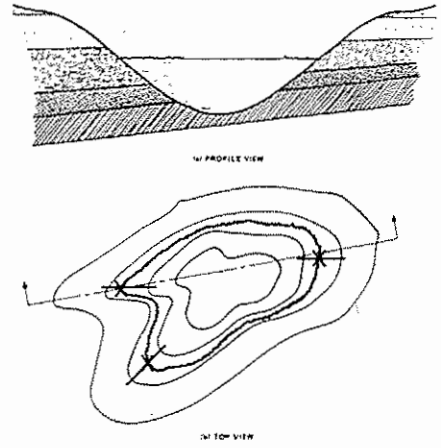


FIGURE 9 RELATIONSHIP OF WATER LEVEL TO TOPOGRAPHY OF TERRAIN

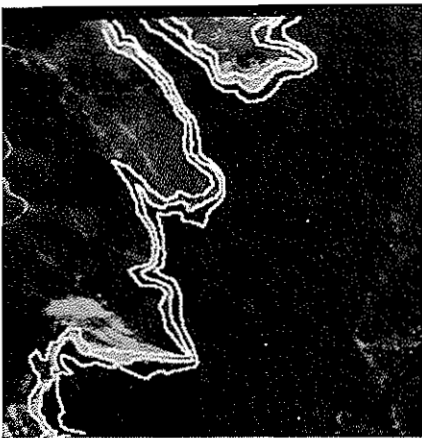


FIGURE 10 TERRAIN CONTOURS SUPERIMPOSED ON IMAGE OF BRIONES RESERVOIR.
The actual water height is 524 feet above sea level.



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

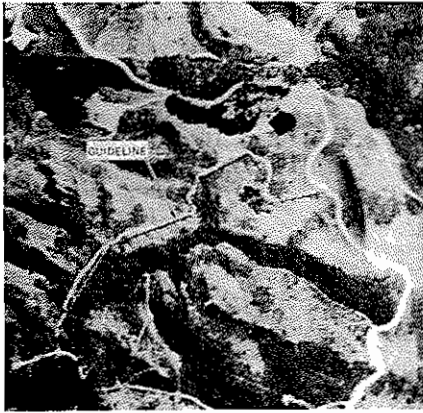


FIGURE 13 A RURAL ROAD WITH GUIDELINE

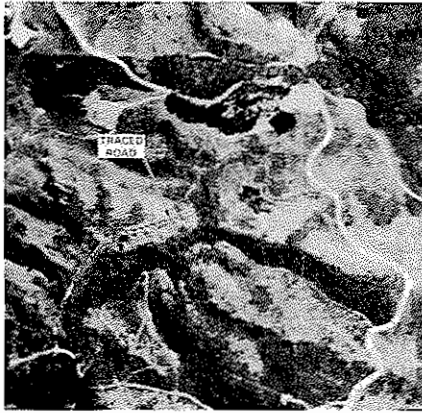


FIGURE 14 OUTPUT OF GUIDED TRACING ALGORITHM

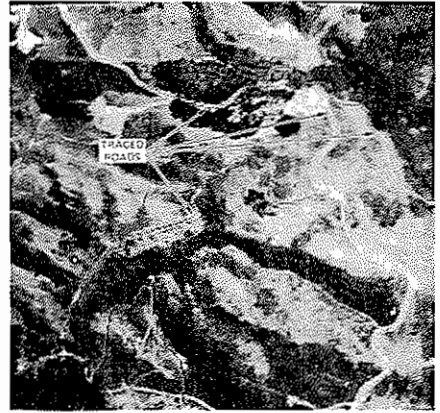


FIGURE 15 GUIDED TRACING OF SEVERAL RURAL ROADS



FIGURE 16 GUIDED TRACING OF SEVERAL URBAN STREETS

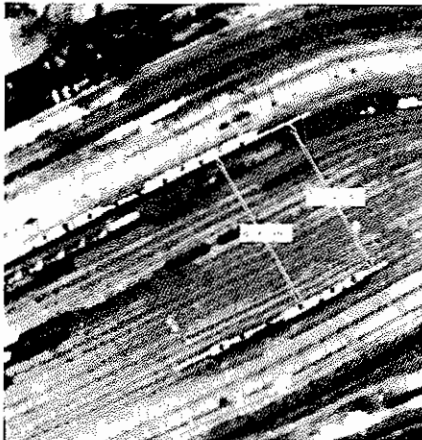


FIGURE 17 AUTOMATED BOXCAR COUNTING

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.

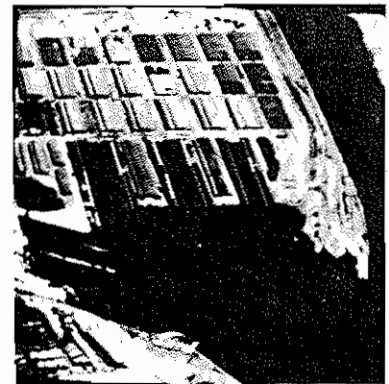
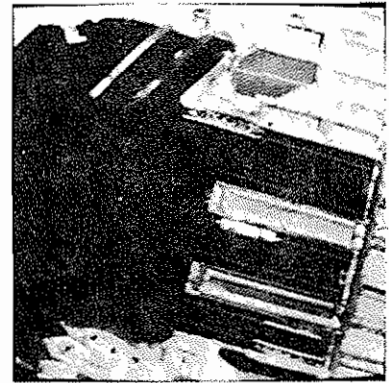


FIGURE 18 AUTOMATIC SHIP MONITORING

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.