

AD-A082 085

RAND CORP SANTA MONICA CA
STRUCTURING THE COMPONENTS OF THE IMAGE MATCHING PROBLEM. (U)
DEC 79 J A RATKOVIC

F/G 17/7

F49620-77-C-0023

UNCLASSIFIED

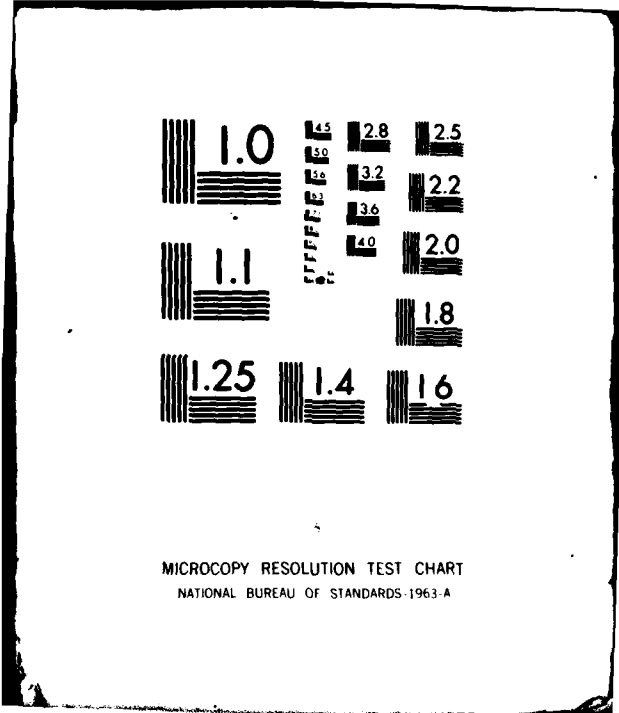
RAND/N-1216-AF

NL

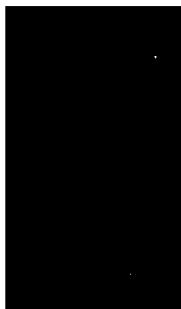
1 1
A
20000

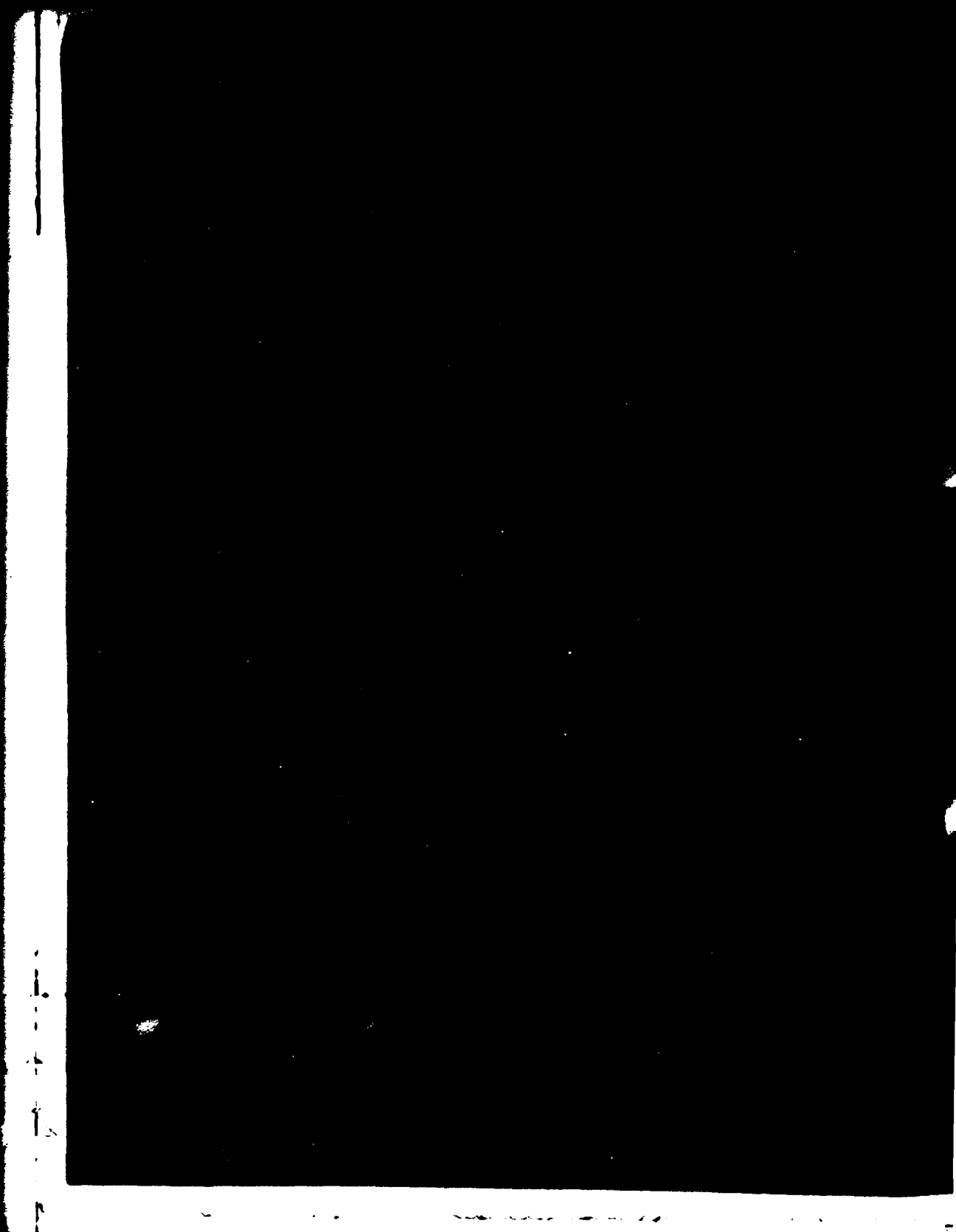


END
DATE
FILMED
4 80
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A






UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)


REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER RAND/N-1216-AF	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Structuring The Components of the Image Matching Problem		5. TYPE OF REPORT & PERIOD COVERED 9 Interim rept.,	
7. AUTHOR(s) J. A./Ratkovic		6. PERFORMING ORG. REPORT NUMBER	
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Rand Corporation 1700 Main Street Santa Monica, California 90401		8. CONTRACT OR GRANT NUMBER(s) 15 F49620-77-C-0023	
11. CONTROLLING OFFICE NAME AND ADDRESS Requirements, Programs & Studies Group (AF/RDQM) Ofc, DCS/R&D & Acquisition Hq USAF, Wash. D.C. 20330		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE 11 December 1979	
		13. NUMBER OF PAGES 27	
		18. SECURITY CLASS. (of this report) UNCLASSIFIED	
		18a. 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) No Restrictions			
18. SUPPLEMENTARY NOTES			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Correlation Sensor Mapping Matching Images			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) See Reverse Side			

DD FORM 1 JAN 73 1473

UNCLASSIFIED 296600
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)



This Note briefly describes the matching process. The major emphasis is in describing the elements of the matching process--the scene, matching algorithms, and errors--and determining their roles in and effect on the matching process. A means is provided for structuring the map matching problem. The scene is defined by the degree of homogeneity and the number of independent elements in each homogeneous region. The errors are further broken up into categories which are mutually exclusive, comprehensive, and positively related to a preprocessing technique or algorithm required to accommodate them. The errors are thus broken up into one of the following categories: global, regional, local, and nonstructured. Finally, the matching algorithms are defined as being of a feature matching correlation or hybrid type. The latter type is a new class of algorithm developed at Rand which bridges the gap between feature matching and correlation types of algorithms. (Author)



12

N-1216-AF

December 1979

STRUCTURING THE COMPONENTS OF THE IMAGE MATCHING PROBLEM

J. A. Ratkovic

DTIC
ELECTE
MAR 19 1980
S D C

A Rand Note
prepared for the
United States Air Force



APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

PREFACE

The accurate guidance of its strategic and conventional cruise missiles is a matter of great concern to the Air Force. There are at present two methods of improving the location accuracy of the vehicle beyond that provided by the onboard inertial system. The first involves time-of-arrival techniques in the Global Positioning System (GPS). Earlier Rand studies of the performance cost and vulnerabilities of this system have shown that a "survivable" GPS system would cost several billion dollars and still may be vulnerable to jamming in the terminal area; thus, terminal delivery in the presence of jamming may not be accurate enough to allow for the delivery of nonnuclear munitions.

The second method, a potentially cheaper alternative to GPS guidance, is correlation guidance. A correlation guidance system using terrain contours (TERCOM) is configured as the heart of the guidance system for the present-generation cruise missile. Eventually, there will be a need for a navigation system that can go anywhere in the world, including to the flat areas where terrain-contour navigation fails, and possibly for the delivery of nonnuclear munitions on both strategic and tactical targets. Correlation guidance schemes using imagery (instead of terrain contours) along the midcourse flight path and in the terminal area are a potential means of achieving these goals.

Current Rand studies are providing a better understanding of the basic principles and limitations of the image-correlation system. They should also provide a methodology for improving the scene selection process and yielding a higher accuracy per fix. Aimed at the problems encountered in using imagery--especially those of radiometrics, in which the Air Force is heavily engaged--this Note is intended to be a first step in providing a unified theory for describing all matching processes (both pattern recognition and correlation) and for understanding the effects of inherent scene characteristics on the performance of the system.

This work was performed under the Project AIR FORCE research project "Battle Management System for ICBMs, Bombers, and Cruise Missiles."

ACKNOWLEDGMENTS

The author would like to thank John Clark and Howland Bailey for constructive discussions during the assembly and structuring of this material. A special note of appreciation goes to Edward Taylor, whose comments improved the quality of the Note and provided useful ideas in understanding the problem. Edward Conrow (of the Aerospace Corporation) and Hyman Shulman also provided valuable thoughts on the subject. Finally, a special thanks to Theodore Garber, who reviewed the Note.

Accession For	
NTIS	GRA&I
DOC TAB	
Unannounced	
Justification	
By	
Distribution/	
Availability Class	
Dist	Available for special
A	

CONTENTS

PREFACE	iii
ACKNOWLEDGMENTS	v
GLOSSARY	ix
Section	
I. INTRODUCTION AND SUMMARY	1
II. DESCRIPTION OF THE MATCHING PROCESS AND ITS COMPONENTS ...	4
The Matching Process	4
The Elements	5
III. STRUCTURING THE PROBLEM	20
Scene Structure	20
Structuring the Errors	24
Preprocessing	25
BIBLIOGRAPHY	27

GLOSSARY

IR	Infrared
LWIR	Long-wavelength infrared
S/N ratio	Signal-to-noise ratio

I. INTRODUCTION AND SUMMARY

The bulk of this Note is divided into two parts. Section II describes the correlation process and its elements, providing a background for structuring the matching process, which is discussed in Sec. III.

This Note describes the structure of a scene in terms of homogeneous regions and discusses general methods for scene decomposition. Four generic types of matching algorithms are discussed--the two basic matching algorithms (image correlation and feature matching) and two variations which merge the correlation and feature-matching processes. Preprocessing is discussed in terms of either compensating for system biases or gain changes or spatial grouping of the elements to compensate for geometric errors.

Finally, the Note discusses four generic classes of error sources associated with the matching process--global, regional, local, and nonstructured errors. It is felt that all errors can be fitted into these mutually exclusive categories and that these categories can be used to uniquely describe the changes in system performance (rather than treating the perturbation in performance due to each error source individually). Figure 1 presents an overview of the entire map matching process in terms of components.

Matching processes can be separated into two phases, as indicated in Fig. 2. Phase one consists of acquisition, where the goal is to avoid a false fix and roughly locate the match position. In phase two different preprocessing and matching algorithms are used to refine the match position to obtain high accuracy.

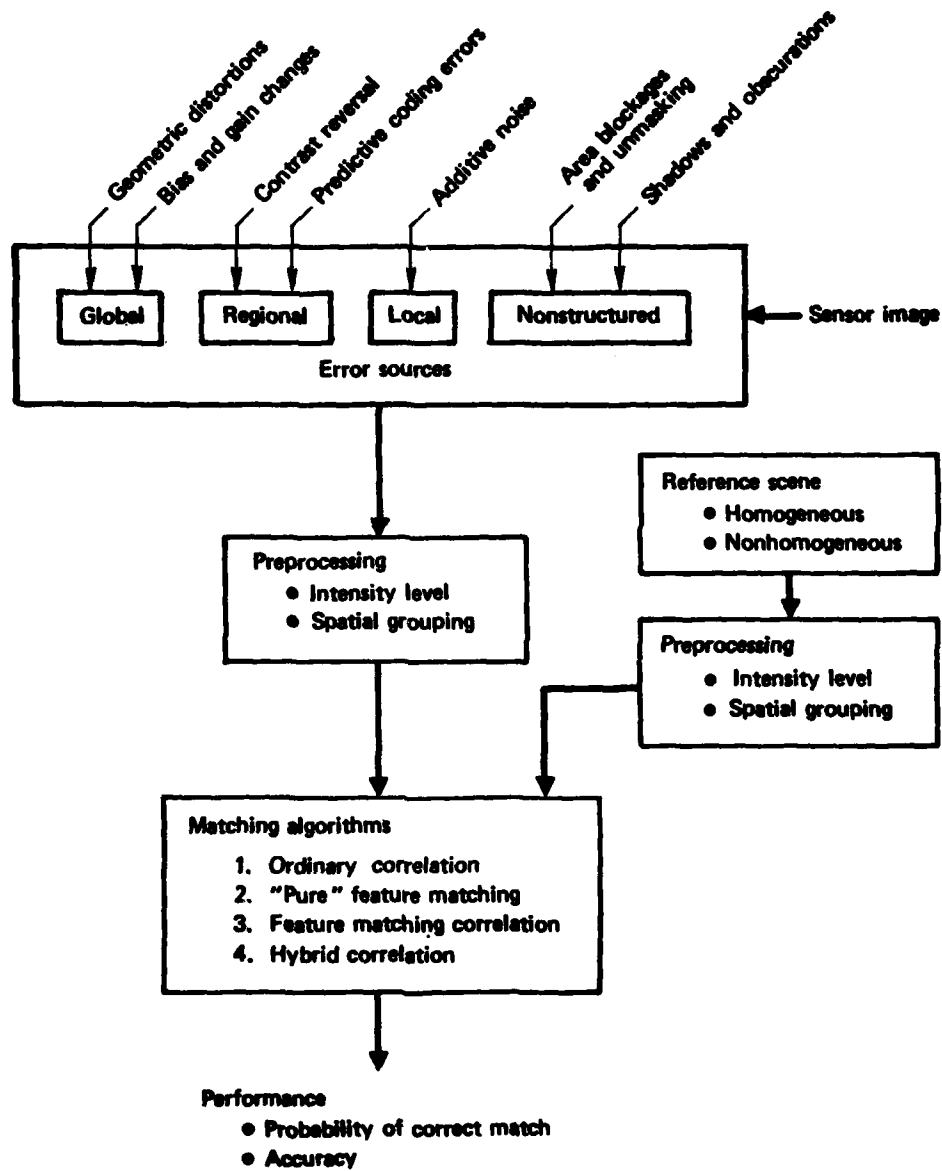


Fig. 1 — Generic overview of map matching process

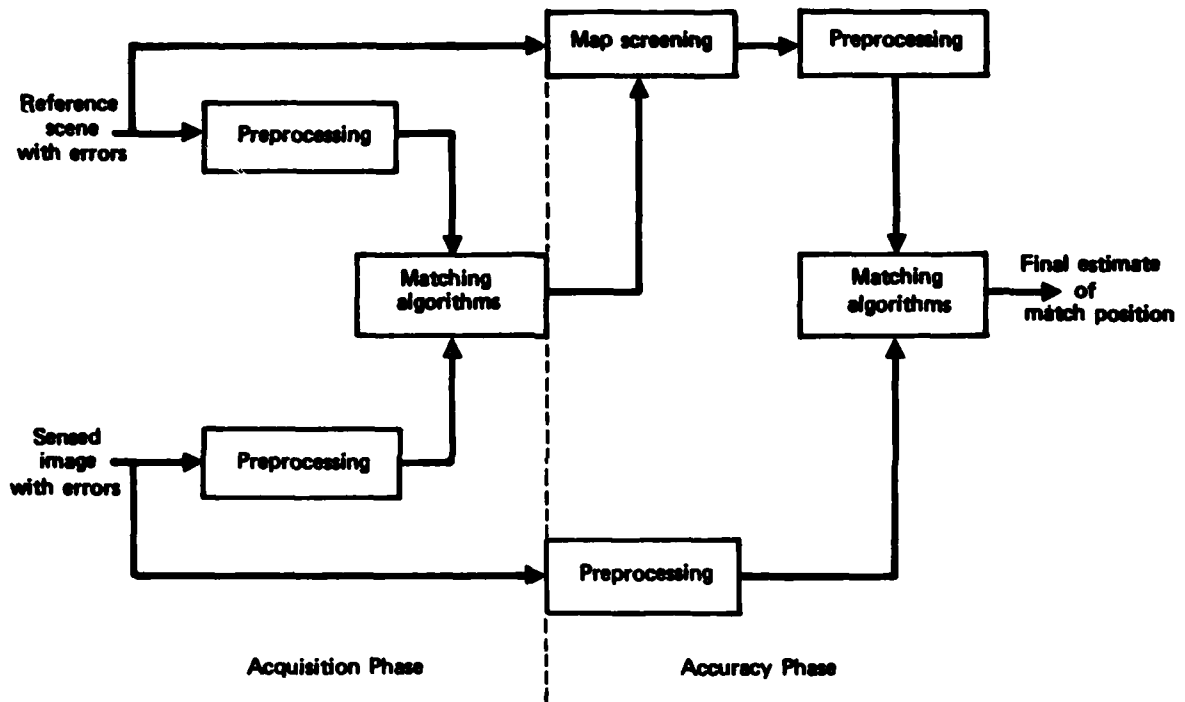


Fig. 2 — Acquisition and accuracy phases of map matching process

II. DESCRIPTION OF THE MATCHING PROCESS AND ITS COMPONENTS

THE MATCHING PROCESS

Figure 3 shows a block diagram overview of the matching process. Here a preselected reference scene or map is chosen which is to be used by a vehicle to make a midcourse or terminal position fix. It is hoped that the reference map size, in combination with the accuracy of the inertial guidance system (updated by a correlation fix at the last check point or initialized at the weapon release point) will be such that the image (or terrain contour in the case of the TERCOM system) taken by the vehicle sensor will fall within its boundary. Comparison of the sensor image with its exact spatial counterpart in the reference map reveals a number of differences. These differences, or errors, exist for a number of reasons. They may be due to changes in the average scene intensity level (in the case of imagery only), sensor noise, problems in the reference map preparation (e.g., incorrect cross wavelength

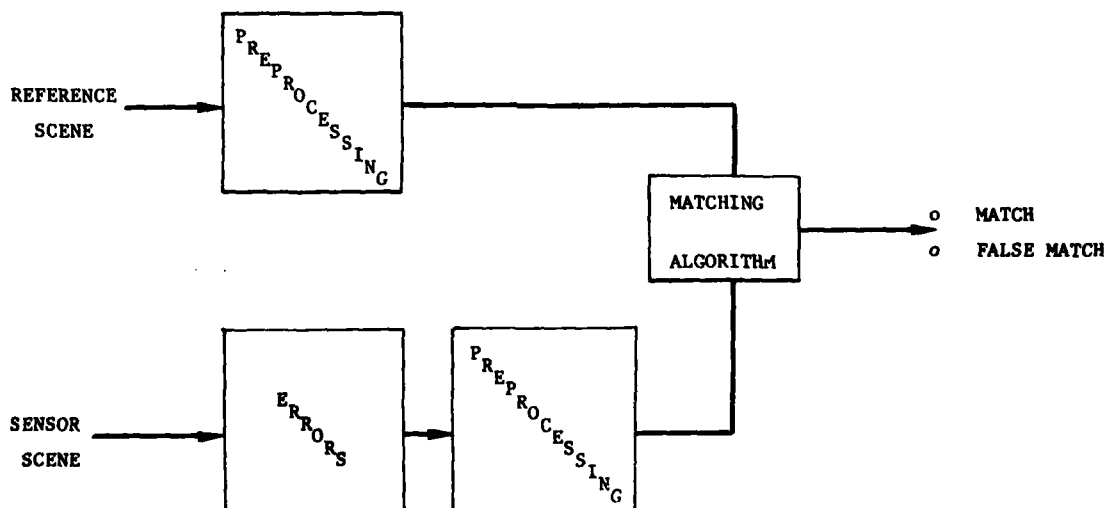


Fig. 3 — Overview of the matching process

predictions when the original imagery used in preparing the reference scene was taken from another portion of the spectrum) or may be due to system errors which cause the sensor to be located at a different spatial point or orientation than originally predicted (causing geometrical distortion between the reference and sensor scenes). Regardless of the exact nature of the errors, from the system point of view all errors can be considered to originate in the sensor image before any other operations are performed on the image.

Both reference and sensor maps can be preprocessed to enhance the ability of the matching algorithm to correctly identify the point at which the sensor image matches the reference scene. The output of the matching process will either be a correct match (performance is measured by the accuracy of the fix) or a false fix (the probability of occurrence is measured).

The basic matching problem can be stated simply as "how does one choose (1) the reference area from the ensemble of possible maps, (2) the preprocessing procedure, and (3) the matching algorithm so as to maximize (either separately or individually) the performance criteria of accuracy and probability of correct match?"

The remainder of this section describes the details of the matching process to obtain a better understanding of the modeling of the process.

THE ELEMENTS

Composition of the Scene

The scene is the most complex component of the map matching problem and the most difficult to model. In the discussion that follows we shall examine "scene composition" (relative to both a visual and a statistical representation of a scene) and methods for decomposing the scene.

Scenes can be described in the visual domain (the eyeball process) as being composed of a set of features. An illustration is the simple scene shown in Fig. 4. Here, for example, the window feature consists of a set of four panes enclosed by a frame.

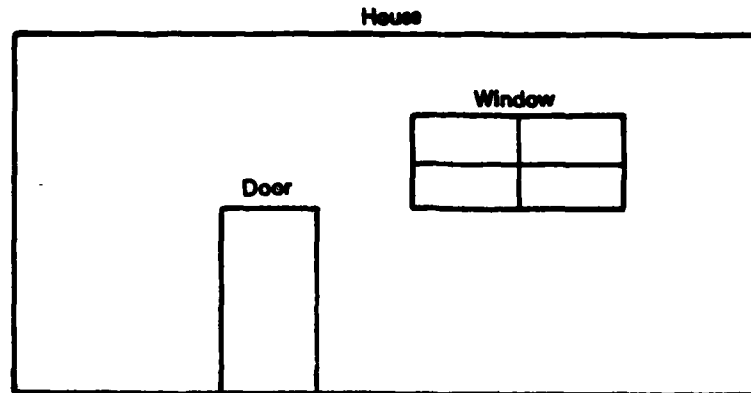


Fig. 4 — Example of features consisting of homogeneous regions

With actual sensor data, picture elements (pixels) are described by a set of intensity values, as indicated in the agricultural scene of Fig. 5. There are regions of intensity values, in the scene which can be considered analogous to features in the visual domain. These are homogeneous regions within the scene. We define a homogeneous region to be a set of spatially connected pixels or elements which possess the statistical property of at least first-order* stationarity and possibly second-order stationarity† and assume that homogeneous regions are equivalent to features (because a feature can be defined by a single homogeneous region or set of homogeneous regions).

In Fig. 5 we have identified four homogeneous regions and tagged each pixel (indicated at the bottom portion of the figure) as belonging to one of the four regions. Examining each region, we see that the intensity value of a given pixel does not vary significantly from the mean value and that there are distinct boundaries (defined by differences in the mean intensity level) between regions.

Thus far we have shown that scenes are composed of homogeneous regions which may be considered equivalent to features. From a physical standpoint, homogeneous regions are areas in which the signature (emissivity for visual and IR, reflectivity for radar, and altitude for

* Mean intensity level constant over the region.

† Mean and variance constant and the autocorrelation independent of position.

terrain contours) is expected to remain fairly constant, e.g., a grassy field in which all the elements in the region are expected to have the same mean value but not necessarily a constant value. (For instance, all the scene elements in a grassy field at LWIR wavelengths are expected to have the same intensity value; however, that intensity value may vary as a function of sun angle, season, etc.). Having established that a scene is composed of homogeneous regions, is there a further subdivision by which we can characterize specific homogeneous regions?

Returning to Fig. 5, we see that there are small variations in the intensity level within a homogeneous region. Some of this variation can be attributed to sensor noise; neglecting this possibility for the moment, however, one can consider the variation to be due to some perturbation in the signature of the region. For instance, one can consider the grassy field not to be uniform, but instead to have a few fallen tree trunks and shrubs dispersed within it. If the ground resolution of the sensor is of the same magnitude as the size of the shrubs and tree trunks, then we would expect variations in the intensity level of the grassy region due to these objects, presuming, of course, that the signature of the objects was different from the grass at the wavelength of the sensor. Thus, we can further categorize a homogeneous region in the physical domain by the number of objects which contribute to a signature variation and in the statistical domain by the number of statistically independent elements* which comprise the region.

The "scene resolution" provides a useful concept in analyzing the statistical variation of a region. We shall define the scene resolution as the number of sensor resolution elements or pixels required to make

* Statistical independence is different from homogeneity. For instance, one can generate a completely random map from a single distribution that will have the property of homogeneity but will also have all the elements independent. One can imagine a homogeneous region containing a number of independent elements, e.g., a desert area in which the shrub patterns (depending on resolution) constitute the independent elements. It is difficult to test for and locate independent elements in a scene. J. A. Ratkovic et al., *Estimation Techniques and Other Work in Image Correlation*, R-2211-AF, September 1977, describes a short-cut method for estimating this parameter by working backwards from the statistics of the correlation surface and assuming a homogeneous scene with all elements independent.

up one independent element in the scene. If there are N pixels within a homogeneous region and N_I independent scene elements ($N_I \leq N$) then the average scene resolution for the region is given by N/N_I . Returning to the grassy field example, if the field were completely uniform with no variations in intensity level, then it could be considered to contain only one independent scene element and the scene resolution would be given by the total number of sensor elements in the region, N . In this particular case, one could not expect to resolve any features within the region due to its uniformity; thus the scene resolution equals the size of the region (in terms of sensor elements). If, however, there were a number of objects (with different signatures) such as tree trunks and shrubs within the grassy region, then we would expect the region to be statistically represented by several independent scene elements. It should be noted also that if the resolution of the sensor were to increase to the point that dimensions of objects within the grassy field covered several sensor resolution elements, then these objects would be considered homogeneous regions in themselves. If the resolution were to increase further, then areas within the objects (e.g., moss on the fallen tree trunks) would eventually become homogeneous regions and the process of identifying homogeneous regions could continue ad infinitum.

At this point we see that for a given sensor resolution it is possible to describe statistically a scene as being composed of a number of independent elements. It will be shown later that the size and number of the homogeneous regions and the constituent number of independent elements in each region play important roles in the matching of scenes.

Decomposition of the Scene

Having described the composition of a scene in terms of homogeneous regions and independent elements within each region, how might we decompose a scene into its fundamental components? The problem can be broken into two subproblems: (1) locating the homogeneous regions and (2) locating the independent elements within a region.

Homogeneous regions can be found visually; however, when one considers the large arrays of numbers involved in describing scenes it is desirable, at the very least, to introduce an automated process to make a first cut at locating homogeneous regions. Automated techniques for locating homogeneous regions can be grouped as being based on edges or on areas. The field of pattern recognition is replete with techniques for locating boundaries within an image based on various forms of edge operators. These techniques apply gradient or Laplacian-type operators to the scene and then use threshold techniques to decide upon the existence of an edge or feature. The major danger in using these techniques is that noise and distortion may make it difficult to locate edges in sensor imagery.

Two area-based techniques can be used to locate homogeneous regions. The property of stationarity of the region can be used to form the basis for separating pixels into regions. In this process, one would attempt to build regions of spatially connected pixels which have the same mean and variance statistics. Another method for screening homogeneous regions would be on the basis of spatial frequency. Returning to Fig. 5, if we were to take a horizontal slice through the data in rows 1 and 11 we would obtain the intensity level plots shown in Fig. 6. As illustrated in the figure, one can associate low spatial frequencies with the homogeneous regions and higher spatial frequencies with the intensity variation within a region. It thus may be possible to locate homogeneous regions by filtering out the higher spatial frequency components.

The number, size, and position of independent scene elements in a homogeneous region can be obtained using recursive image partitioning algorithms. These techniques attempt to group pixels into blocks such that the mutual information between them measured by entropy is minimal. If only the number of independent elements in a homogeneous region is desired, then this can be rapidly estimated using a "statistical scene model" approach. The basic idea is to model the statistics of the correlation surface based on all the scene sensor elements being independent and then working backwards from the correlation statistics to the number of independent elements in the scene.

Agricultural Intensity Values
(Region 12)

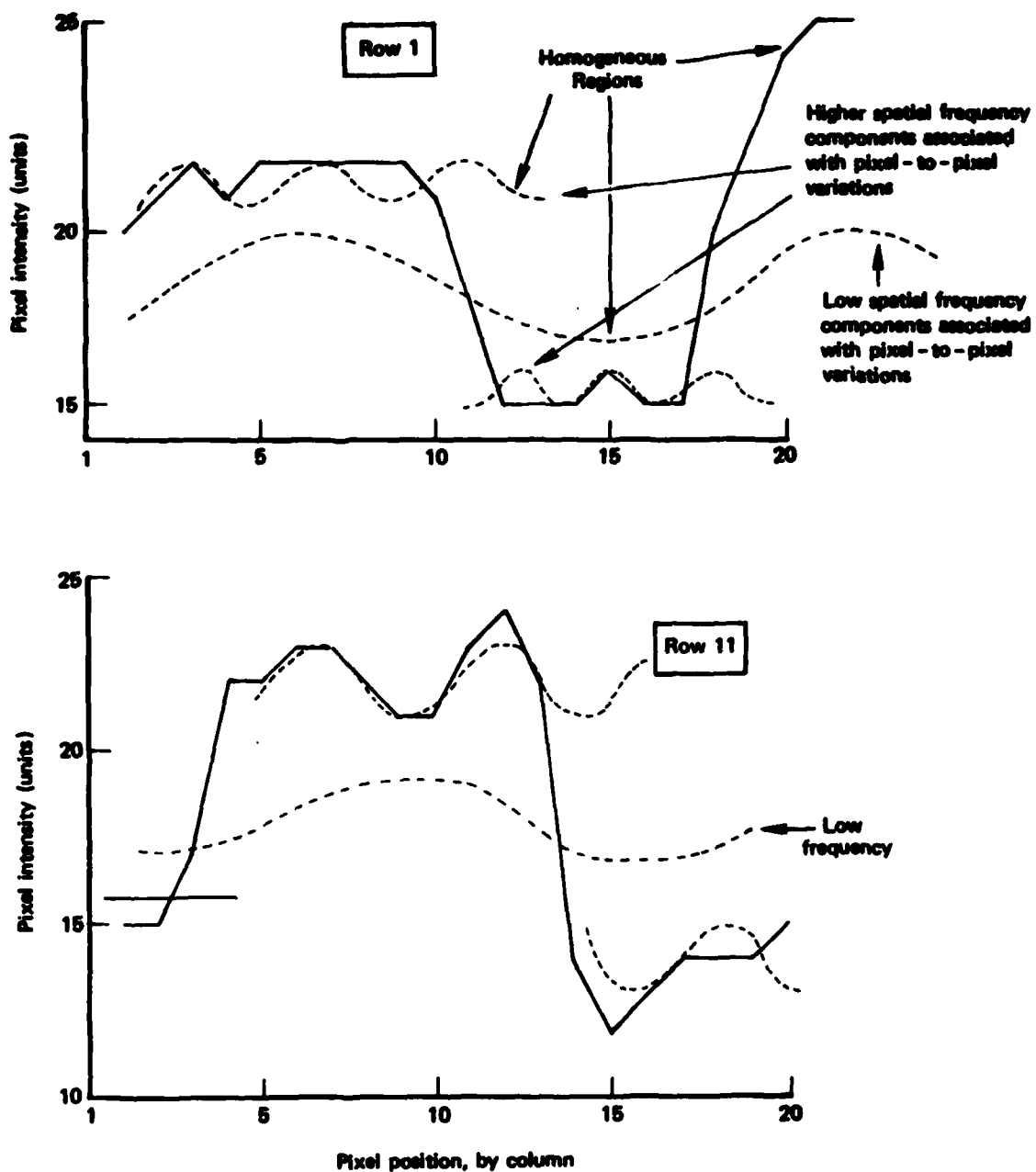


Fig. 6 — Cross-sectional views of map intensity level data for Region 12

Matching Algorithms

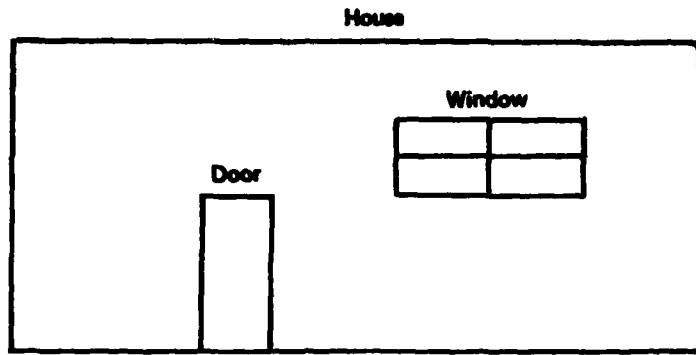
The basic matching algorithms belong to a feature matching or to an image correlation class of algorithms. None of these algorithms have been mathematically derived to maximize system performance (probability of correct match or accuracy) and, therefore, must be considered to be "ad hoc." A subsequent Note* will discuss the development of an optimal algorithm. There are two reasons for presenting these matching algorithms even though they are "ad hoc" and not "optimal." First, they serve to acquaint the reader with the generic types of algorithms being pursued. Second, the optimal algorithm may either be too difficult to implement, in which case the present set of algorithms will provide a fallback position, or (as might be suspected) the optimal algorithm may reduce under a certain set of conditions to a form similar to simple correlation algorithms.

All algorithms basically perform three operations: (1) the establishment of a metric, (2) the computation of that metric for all possible positions of comparison between the reference and sensor maps, and (3) a selection rule for delineating the match position based on the metric value.

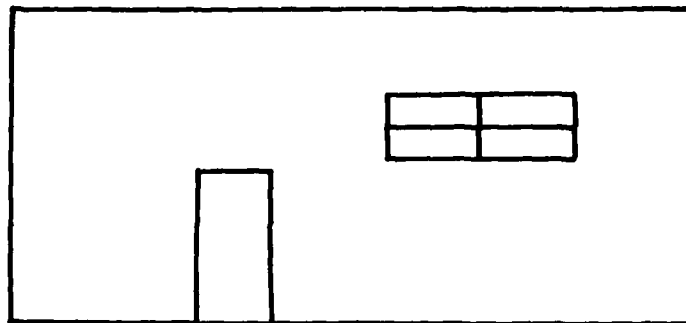
Before these operations can take place, it is first necessary for the "feature matching" procedure to extract the features from the scene. Figure 7 shows a generic description of the process for the simple house scene shown in Fig. 4. The first part of the feature extraction process involves locating the edges or boundaries of features. As indicated in Fig. 7, the scene can be reduced to a set of lines which are the boundaries of the feature. Next the line intersection points are located, as shown in Fig. 7. In general, the number of lines emanating from each vertex is retained and used as part of the weighting criteria in the feature matching algorithm.

In image correlation there are two basic types of algorithms--those that emphasize the degree of similarity between scenes, such as the product, and those which emphasize differences between scenes, such as the difference squared and MAD (Mean Absolute Difference) algorithms.

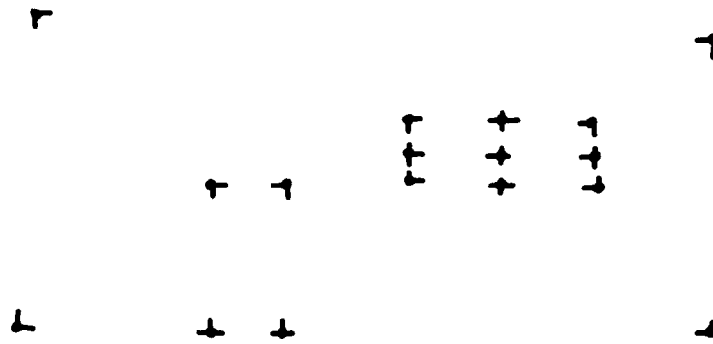
* J. A. Ratkovic, *Performance Considerations for Image Matching Systems*, N-1217-AF, December 1979.



(a) Original scene from Fig. 4



(b) Scene with lines extracted



(c) Line intersection points within scene

Fig. 7 — Generic description of feature extraction technique

To explain the matching process further, it is necessary to make a few definitions concerning the map. First, as shown in Fig. 8, it is explicitly assumed that the sensor map is smaller than the reference map and that the intensity level of an arbitrary sensor element is Y_I , whereas that of the reference map is X_I . The displacement of the sensor map from the correct location is the displacement vector, J . In the absence of geometric errors, all elements of the sensor map are congruently positioned with the corresponding elements of the reference map when the displacement vector is zero. At a displaced map position an arbitrary sensor map element, Y_I , is compared to an arbitrary reference map element, X_{I+J} . If the sensor map contains N elements, the most commonly used correlation metrics can be expressed as

$$\phi_{\text{PROD}}(J) = \frac{1}{N} \sum_{I=1}^N X_{I+J} Y_I \quad (\text{Product})$$

$$\phi_{\text{DS}}(J) = \frac{1}{N} \sum_{I=1}^N (X_{I+J} - Y_I)^2 \quad (\text{Difference-Squared})$$

$$\phi_{\text{MAD}}(J) = \frac{1}{N} \sum_{I=1}^N |X_{I+J} - Y_I| \quad (\text{Mean Absolute Difference})$$

In attempting to properly locate the sensor map relative to the reference map, we must compare the sensor map with equally sized portions of the reference map at all possible displacement positions within the reference map boundary. At each point of comparison or displacement position, J , a value of the metric is computed. The selection rule for picking the correlation value associated with the correct match position is to select the extremum value. In the case of the product

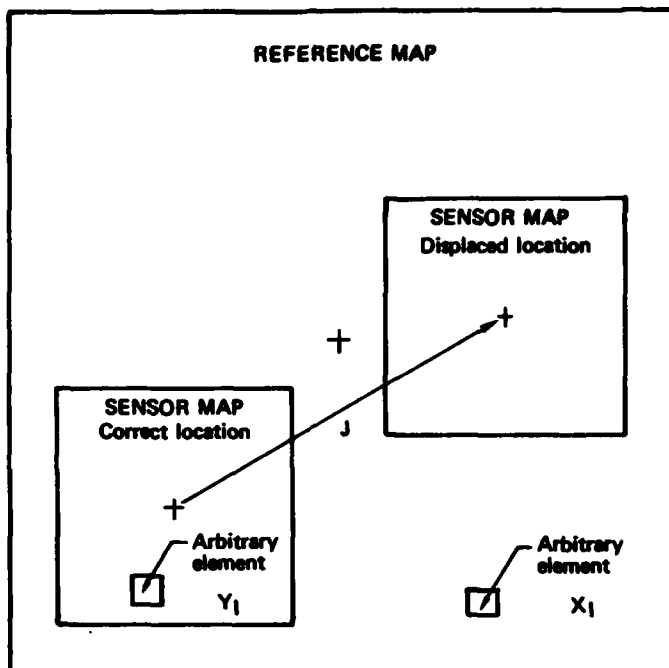


Fig. 8 — Map definitions

metric, the correlation value should be maximum at the correct match point; whereas the difference-squared and MAD metrics should be at a minimum at the correct match point.

As pointed out previously, feature matching algorithms do not use the intensity levels of the scene but generally start with a transformed map which contains only the vertices of line intersections within the scene, as illustrated in Fig. 7. Having transformed the map to vertex data, the feature matching algorithms take on the appearance of a weighted difference-squared algorithm. They proceed by placing the sensor map at a specific reference map vertex. These algorithms then measure the difference in position between all other points in the sensor map and the closest points in the reference map. The metric then proceeds to sum up all of the position differences (generally weighted by the number of line intersections associated with the point) by a weighted least-squares or difference-squared-type algorithm. This metric is then computed for all possible positions of the transformed sensor map within the reference map boundary and the minimum value of the metric is chosen as the position of best fit between the two maps.

Errors

There are a number of error sources that can degrade the performance of map matching systems. These include: (1) geometrical distortion, (2) bias and gain changes in the scene intensity level, (3) region level intensity shifts, (4) area blockages, (5) additive noise, and (6) predictive coding errors. These errors are described briefly below.

Geometric distortion of the sensor map coordinates relative to the reference map coordinates degrades, in ways that are discussed below, the performance of a map matching system. The four most important types of geometrical distortion are errors in synchronization, rotation, scale factor (magnification), and perspective. The detailed analysis of these effects, for digital systems, involves synthesizing a grid of cells each of which is given a value that is an appropriately weighted average of the values of the distorted cells that partially overlap each of the undistorted cells. These errors are illustrated in Fig. 9 where, for each case, the four cells surrounding the center of the reference map are depicted, together with the corresponding cells of the distorted sensor map.

Synchronization errors occur because there is no way to ensure a common origin between the sensor and the reference map grids. As shown in the figure, this type of error results in all the grid elements of one map being fractionally displaced from those of the other map. This displacement can cause each sensor map grid element to overlap as many as four grid elements of the reference map. The effects of synchronization errors are most significant when the dimensions of a sensor element are comparable to the average dimensions of a statistically independent scene element.

Rotation errors can be caused by heading or attitude reference errors on board the vehicle. If the sensor map is centered but rotated relative to the reference map, the map matching process compares a single sensor cell with a combination of fractions of both matching and nonmatching reference cells. The amount of overlap with nonmatching cells increases as one moves radially outward from the center of the two maps.

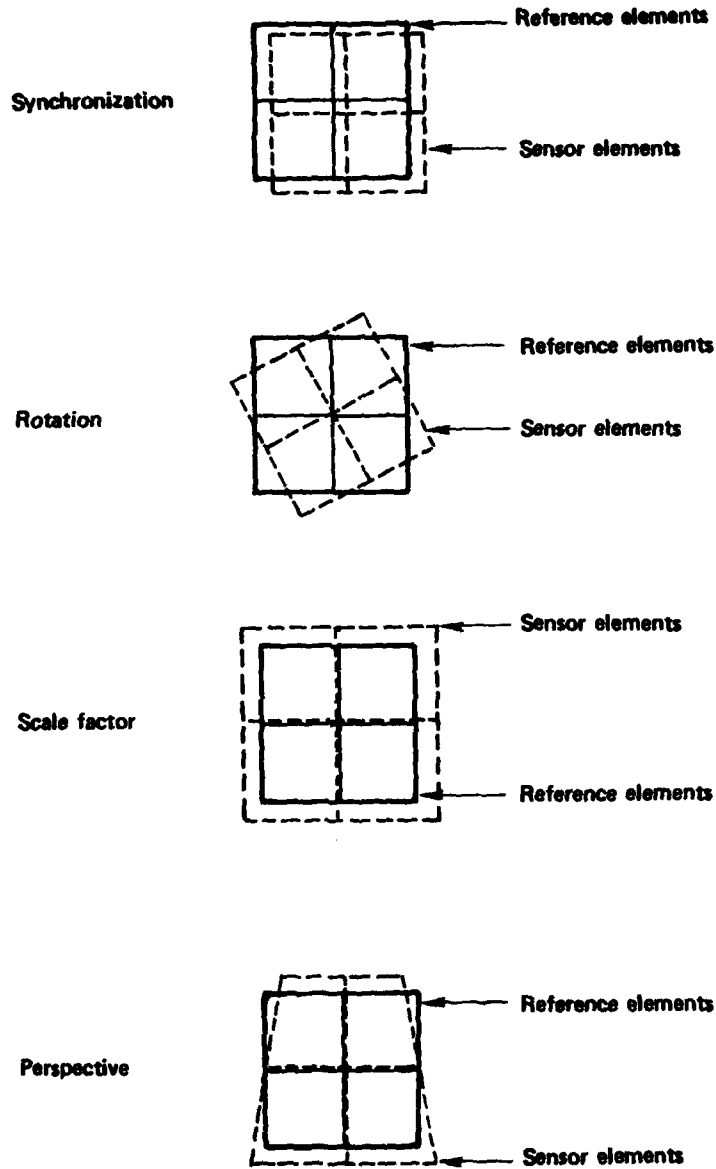


Fig. 9 — Geometric distortion errors

Uniform magnification or scale errors are primarily caused by errors in altitude or range to the target, although in some cases they may be caused by several other effects as well. In the presence of scale factor errors, the sensor elements are dimensioned either somewhat larger or somewhat smaller than the reference map elements. Consequently, elements of the sensor map, when overlaid on the reference scene, will encompass both matching and nonmatching reference elements, with the amount of nonmatching overlap increasing as one moves radially outward from the center.

Perspective errors occur when the sensor views the reference area from a different position in space, because of midcourse navigation inaccuracies, for example. Owing to the difference in perspective, a grid pattern of square cells is transformed into an array of trapezoids. Thus, the effect is similar to a linearly varying scale factor error.

When geometrical distortions are present, only a partial match between sensor and reference map elements is possible. When the map centers are slightly displaced, some of the previously nonmatching map elements are brought into coincidence, so that a partial match condition holds for these displacements. The overall effects on the correlation function or comparison metric are thus twofold: the peak value of the metric for the matched condition is reduced and the breadth of the function is increased.

The sensor may introduce both *bias level* and *gain changes* throughout the entire scene. If the scene itself has a great deal of intensity level variation, it may be difficult to assess whether (a) a gain or bias change has occurred, or (b) the sensor has imaged an area of the reference map where those intensity levels are present.

As described earlier in this section, a scene is composed of a number of homogeneous regions. In the case of imagery, as opposed to terrain contour data, the *intensity level of regions may shift* due to sun angle, seasonal or atmospheric effects, etc. In processing the scene, one should be aware that the region levels may shift in mean value relative to one another.

Uniform amplitude errors affecting contiguous areas are referred to as block substitution errors or *area blockages*. Shadows due to

scattered low clouds or changes in sun angle can cause dark blocks and intervening sunlit clouds and certain kinds of jamming can produce bright blocks. Errors of this sort can generally be categorized by the amplitude level and size of the area affected.

Additive noise can be either constant in value over the scene or multiplicative with the amplitude value dependent on the scene level. It can generally be categorized by its frequency spectrum and the S/N ratio.

Predictive errors arise when the reference map must be created synthetically from original imagery at a different wavelength and possibly at a different aspect angle. To estimate the signature of the imagery, it is necessary to determine the physical attributes (e.g., material content) of the scene being imaged and develop a three-dimensional geometrical reconstruction program from which to estimate the signature. Errors arise from either an inability to correctly estimate the signature associated with a scene (because no reference data are available at the same wavelength) or from the use of an average signature. In the latter case, some sensor wavelengths may have scene signatures which are highly time varying, and to avoid modeling the signature for the exact moment of arrival of the vehicle over the target area, an average signature may be used. Generally these errors are regional in nature, i.e., homogeneous regions are modeled with the wrong mean level and variation.

This section of the Note has described the overall correlation process and discussed each of the elements of the process--scene, decomposition, matching algorithms, and errors. In the next section of the Note we discuss optimal and suboptimal performance measures by which to judge the process.

III. STRUCTURING THE PROBLEM

The first part of this Note indicated that map matching is complex, involving a large number of error sources, numerous types of matching algorithms, and a scene which is difficult to model. Rather than dealing with an almost endless list of errors, scenes, algorithms, and preprocessing, it would be desirable to develop a generic structure for the problem, with each category in the structure directly linked to an effect on system performance. This section is designed to (1) reduce the number of components of the map matching process by providing a generic categorization of these components, and (2) through the use of this categorization, provide an overall framework for the problem for simplification. A subsequent Note will use this structure to explain the effects of algorithms, scenes, preprocessing, and errors on system performance (accuracy and probability of false match).

SCENE STRUCTURE

As described previously, the scene can be described statistically as being either homogeneous or nonhomogeneous: practically all real-world scenes are nonhomogeneous and are thus described by the size and number of homogeneous regions within the scene and the interpixel correlation between adjacent pixels within a region. Both area-based methods (using the statistical properties of the scene) and edge-based methods (using the gradients between boundaries of features) can be used to decompose the scene into a set of features or homogeneous regions.

Correlation Methods

The standard correlation process works on the gross characteristics of the scene and all preprocessing is done globally (i.e., the mean level when subtracted out is zero-measured over the entire scene, and, similarly, when the scene is normalized by the variance this is done over the entire scene). In a sense, the usual correlation process is designed to work on a homogeneous scene. There are two basic

variations to the standard or usual correlation algorithm which are more specifically tailored to nonhomogeneous scenes and the errors associated with them. It should be noted that these variations, in the absence of nonhomogeneity in the scene, reduce to the usual correlation process. We denote these variations that deal with scene nonhomogeneities as (1) feature matching and (2) hybrid algorithms.

One could introduce a feature matching algorithm into the correlation process by breaking up the sensor and reference maps into homogeneous subareas. Each of these maps would then consist of a set of homogeneous regions and all processing (rather than being on a global scale) would then be performed separately on each homogeneous subregion. Thus, when maps are zero-meanded and normalized, the local mean and variance in each subregion can be computed and used to perform the normalization.

After processing both the reference and sensor maps on the basis of homogeneous regions, a standard correlation algorithm can be used to determine the position of match between the two maps. The major generic difference between this feature matching correlation algorithm and the "pure" feature matching algorithm (employing pattern recognition techniques) is the weighting given to homogeneous regions. In "pure" pattern recognition algorithms, edges are first extracted and used to identify line intersection points. These line intersection points or vertices then form the primary basis for matching two scenes. In a sense (since edges can be considered the boundaries of homogeneous regions, and vertices are formed by the intersection of edges) a "pure" feature or pattern matching algorithm weights all homogeneous regions equally, whereas in the feature matching correlation algorithm, each homogeneous region would receive a weighting proportional to its size (measured in terms of the number of independent elements contained within). In summary, "pure" feature matching algorithms can be viewed as being different from feature matching correlations in that different weights are assigned to the various homogeneous regions.

There is another adaptation of the standard correlation algorithm which has been developed at Rand that one can implement to accommodate homogeneous regions. We shall refer to this as a hybrid algorithm

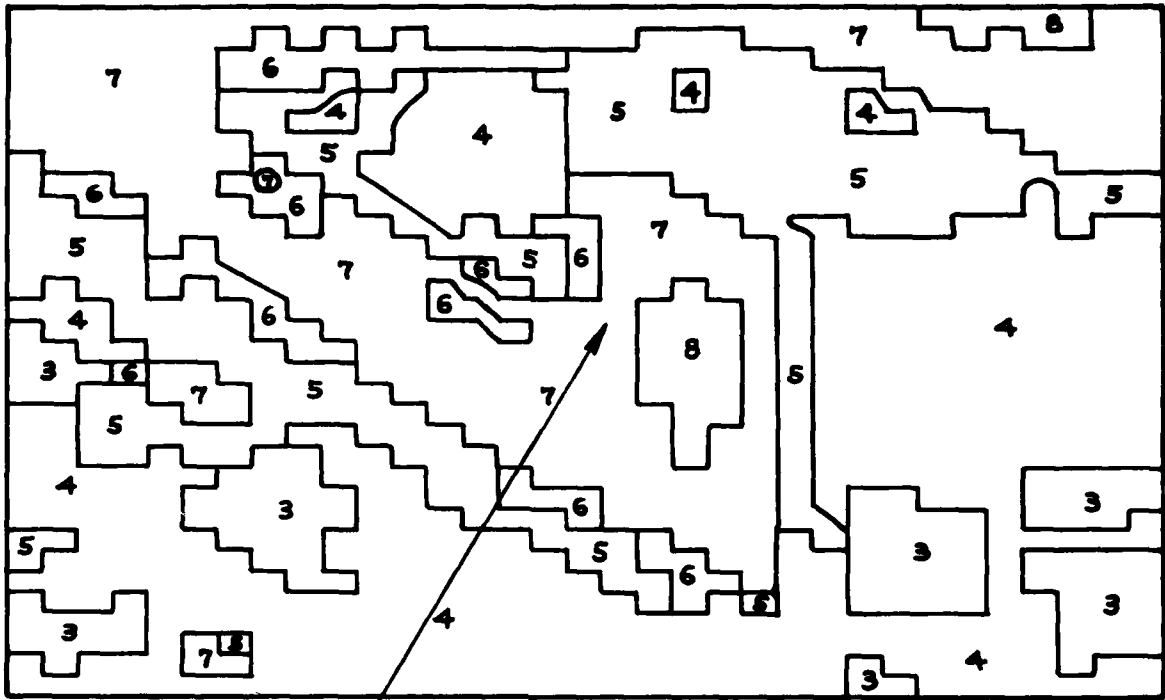
which processes only the reference scene into homogeneous regions. The principal idea here is that every position of comparison between the two images is assumed to be the correct one. Thus at each displacement position or comparison point the sensor scene is segmented identically as its counterpart reference map.* At the position at which the two maps correctly match, the sensor scene will then be segmented almost perfectly, enhancing the match, and at all other positions the sensor map segmentation will essentially look like noise. The objective of this correlation method is to avoid the errors associated with extracting homogeneous regions or features from the sensor image and the additional processing requirements placed on the system. If the image is noisy, normal edge operators have difficulty in performing their feature extraction task and, as a compromise, the hybrid approach, which strictly is not as good as a "pure" feature matching or correlation feature matching algorithm, does possess significant advantages over the standard correlation approach at accommodating certain types of feature errors such as contrast reversals.

In Fig. 10 we show an example of this hybrid processing scheme. We have in the figure identified each reference pixel with a homogeneous region. Thus each reference pixel has both a region identification and an intensity associated with it. The template for the sensor map processing is shown for two map displacement positions. As indicated in the figure, the sensor map is segmented into homogeneous regions at each of these displacement positions in a manner identical to that of the reference map elements occupying the same spatial position. The sensor map elements are then processed by homogeneous regions (i.e., the mean intensity level subtracted out and possibly normalized by the intensity variation in the region) with the total correlation between sensed images and reference map being the sum of the correlation in each region at each displacement position.

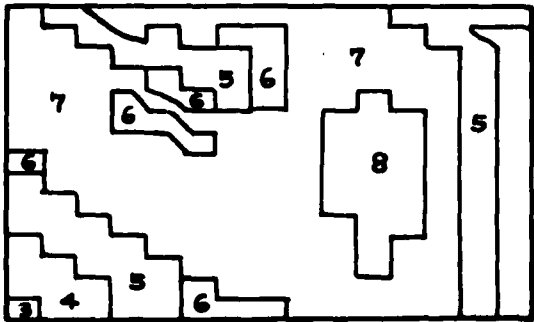
We have identified four generic types of image matching methods:

* For each displacement position the matching process consists of correlating each homogeneous region of the reference map and segmented sensor image separately, and combining additively the correlation in each individual region.

Reference map segmented into homogeneous regions



Template for sensor map processing for center displacement position



Template for sensor map processing for last displacement position

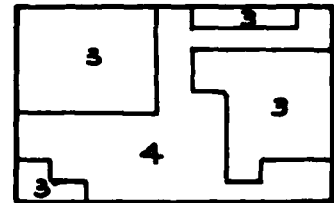


Fig. 10 — Illustration of hybrid matching process

1. Standard correlation algorithm
2. "Pure" feature matching algorithm
3. Feature matching correlation algorithm
4. Hybrid algorithm

The first two methods are the two basic approaches to image matching while the latter two methods are variations of the standard correlation process designed specifically to accommodate nonhomogeneous scenes and the nonglobal errors associated with them.

STRUCTURING THE ERRORS

There are a number of error sources, as indicated above, that affect the performance of the system. It is desirable to lump these errors into generic categories in discussing system performance rather than treating each error source separately. Such a generic categorization should possess the following properties:

1. The error categories should be mutually exclusive.
2. They should be comprehensive.
3. There should be a positive relationship between the category and a specific preprocessing technique or correlation algorithm to accommodate all errors in that category.

Based on the types of errors that occur in the map matching process and the statistical description of the scene, the following generic categories of errors are proposed:

1. *Global Errors*--those errors which uniformly affect equally the intensity level of all scene elements. This category would include geometric distortions and bias and gain changes.
2. *Regional Errors*--those errors where the change in intensity levels occurs uniformly only within homogeneous regions or features within the scene. Examples would be region-level shifts (contrast reversals) and predictive coding errors.

3. *Local Errors*--errors expected to affect each pixel or grouping of pixels (contained within an inter-pixel correlation length) independently. The primary example of this error source is additive noise.
4. *Nonstructured Errors*--this is a rather catchall category designed to fit those errors whose effect on the scene cannot be described as being global, regional, or local (an example of this catchall category is when a cloud cover over the target area casts a ground shadow which changes the signature in a nonstructured manner).

Although some errors may sometimes fit into more than one category, this generic categorization will normally accommodate all error sources as well as provide a convenient means of establishing guidelines for algorithm and preprocessing selection.

PREPROCESSING

The preprocessing of sensor imagery consists of either changing the intensity levels through the image or segmenting the scene spatially into groups of pixels. The intensity level preprocessing is designed to compensate for any biases or gain changes in the system; spatially grouping of elements is designed to accommodate geometric errors.

In general, preprocessing is designed to accommodate global errors that occur in the scene and which, by definition, affect all scene elements equally. Thus global errors such as gain changes and bias errors are handled by normalizing the intensity level and by zero meaning the data, respectively. As discussed previously, geometric errors also are global in nature and reduce the degree of congruence between sensed image and reference image. To reduce the effect on system performance, geometric errors always force one to work with smaller map sizes and, depending on the nature of the distortion (in azimuth and elevation), may also force one to shape the window of the sensed image. Thus, to accommodate this type of error, it is necessary at a minimum to spatially group the sensor map elements into a single (or number of) smaller map(s). If distortions are uneven in azimuth and elevation it

will also be necessary to spatially group the elements so that the appropriate window shape may be obtained. The reference map will or will not be segmented into features or homogeneous regions depending on whether a feature matching class of algorithm is used.

BIBLIOGRAPHY

- Bailey, H. H., et al., *Image Correlation, Part I, Simulation and Analysis*, The Rand Corporation, R-2057/1-PR, November 1976.
- Brice, C. R., and C. L. Fennema, "Scene Analysis Using Regions," *Artificial Intelligence*, Vol. 1, 1970, pp. 205-226.
- Davis, L. S., "A Survey of Edge Detection Techniques," *Computer Graphics and Image Processing*, Vol. 4, September 1975, pp. 248-270.
- Farag, R.F.H., "An Information Theoretic Approach to Image Partitioning," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-8, No. 11, November 1978.
- Garnett, J. M., III, and S. S. Yaw, "Nonparametric Estimation of the Bayes Error of Feature Extractors Using Ordered Nearest Neighbor Sets," *IEEE Transactions, Computers*, Vol. C-26, January 1977, pp. 46-54.
- Gupta, J. N., and P. A. Wintz, *Closed Boundary Feature Finding Selection and Classification Approach to Multi-Image Modeling*, Laboratory for Applications of Remote Sensing, Note No. 062773, Purdue University, W. Lafayette, Indiana, 1973.
- Kettig, R. L., and D. A. Landgrebe, *Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects*, Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, Indiana, 1975.
- Lahart, M. J., "Local Segmentation of Noisy Images," *Optical Engineering*, Vol. 18, No. 1, pp. 76-78, January-February 1979.
- Pratt, W. R., *Digital Image Processing*, Wiley & Sons, New York, 1978.
- Robertson, T. F., et al., *Multispectral Image Partitioning*, Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, Indiana, 1973.
- Rosenfeld, A., and A. C. Kak, *Digital Picture Processing*, Academic Press, 1978.
- Wacker, A. G., *A Cluster Approach to Finding Spatial Boundaries in Multispectral Imagery*, Laboratory for Agricultural Remote Sensing, Purdue University, W. Lafayette, Indiana, 1970.