An Operator-Based Matching System

Joshua S. Greenfeld
Department of Civil & Environmental Engineering, New Jersey Institute of Technology, Newark, NJ 07102

ABSTRACT: A stereo matching system based on mimicking the operation of a human photogrammetric operator is presented. In the proposed system, stereo matching is viewed as an intelligent vision problem rather than merely an image matching process.

A simplified version of the conceptual model for automating the photogrammetric operator was implemented and tested. The goal of this implementation was to perform a relative orientation without any human operator guidance. Results show that an arbitrary near vertical (not oblique) stereo pair can be successfully oriented to within 0.15 pixel without *a priori* assumption on the relative alignment of the corresponding images (except for the approximate percentage of the overlap).

INTRODUCTION

THE TREMENDOUS ADVANCEMENT in computing power, microprocessing, and software architecture resulted in significant enhancements in the photogrammetric process. Photogrammetry today is more accurate, more versatile, more user friendly, and less time consuming then ever before. Its application fields have been broadened from traditional topographic mapping to a precise measuring device in a wide range of applications ranging from geodetic control to industrial tools. One aspect of the photogrammetric process which made only moderate progress over the last three decades is automatic stereo matching. Most of the problems experienced in the early prototypes of automatic stereoplotters still hamper the realization of a fully automatic system. Most current matching systems presume that there exists an approximate match between the images, and their task is to refine the match. If the system "loses lock" on corresponding images, a human operator must re-establish a coarse match between corresponding image patches.

Traditionally, in photogrammetry, area correlation methods have been employed to perform automatic stereo matching. In recent years, however, the method of interest operator has gained popularity mainly because it renders accurate point positioning. This method is like a pointing refinement tool for accurately matching distinct point-features extracted from closely matched image patches. Both solutions have been applied with limited success, but failed to provide a global (comprehensive) matching solution.

In this paper, we introduce a matching system which is based on the idea of mimicking the best known working matching system, namely, the human photogrammetric operator. Instead of interpreting the matching problem as a problem to find "similar images," we suggest that it be perceived as an intelligent stereo vision problem. In spite of Julesz's (Julesz, 1960) proof (using random-dot stereograms) that stereo itself is not necessarily an intelligent process, we believe that some matching problems, such as discriminating ground from tree tops, require access to knowledge about the spatial arrangement of objects in space.

In the next section we define the stereo matching problem. The definition is necessary, because a correct definition and understanding of the basic problem is essential for developing a robust solution. Then we describe the major factors one has to consider in designing an appropriate matching system. The human operator aspects of stereo matching are then discussed. Our approach to a comprehensive matching system is presented. Some initial experiences and results are presented. Finally, concluding remarks are given.

DEFINITION OF STEREO MATCHING

Stereo matching in photogrammetry is usually performed by analyzing (correlating) the gray level content of two image patches (windows) in a very local domain. These patches are assumed to be (a) coarsely matched, (b) owning similar radiometric properties, (c) portraying the objects in a similar geometric relation (no distortions), and it is assumed that (d) the texture of the various components of the image is distinctive. The exact technical details on how closely the images have to be matched, and what constitutes similar images and measures for similarity, vary from one solution to another. Departures from these restricting conditions bring most systems to a halt and a human operator is required to resolve the match. *A priori* assumptions and restrictions are exercised in both digital area correlation and interest points matching.

In computer vision, the stereo matching process is frequently referred to as the process of solving the *correspondence problem*. It is defined as

> Identifying features (points, lines or structures) on the left and right images which correspond to the same physical object in the scene (Grimson, 1981).

Accordingly, the solution for the correspondence problem does not imply matching of merely two similar images, but rather matching two representations of the same object. Matching is performed in context rather than in an arbitrary quantitative domain. Feature matching is rather independent of their quantitative representation (gray levels) in the image space. Matching edges (boundaries of an object), for instance, qualify as a valid solution for the corresponding problem as the edges represent physical objects in the scene. Figure 1 illustrates the difference between image-based (using gray level values) and feature-based (using an edge) matching.

Marr, in his book *Vision*, uses a rather blunt statement on this matter:

> The job of stereo fusion is to match items that have definite physical correlates, because the law of physics can guarantee only that items will be matchable if they correspond to things in space that have a well-defined physical location. Gray-level pixel values do not. Hence, gray-level correlation fails. (Marr, 1982, page 75).

The notion of matching in the object domain (groundel) rather than in the image domain (pixel) has been recently introduced by Helava (1988). His approach is based on the assumption that image densities, corresponding to each groundel, can be analytically computed if pertinent geometric and radiometric parameters are known. The matching is performed between the predicted image density and the recorded one. The uncertainty in the parameters of a particular groundel are solved by least squares. This solution relies on our ability to model correctly

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING,
Vol. 57, No. 8, August 1991, pp. 1049–1055.

0099-1112/91/5708–1049$03.00/0

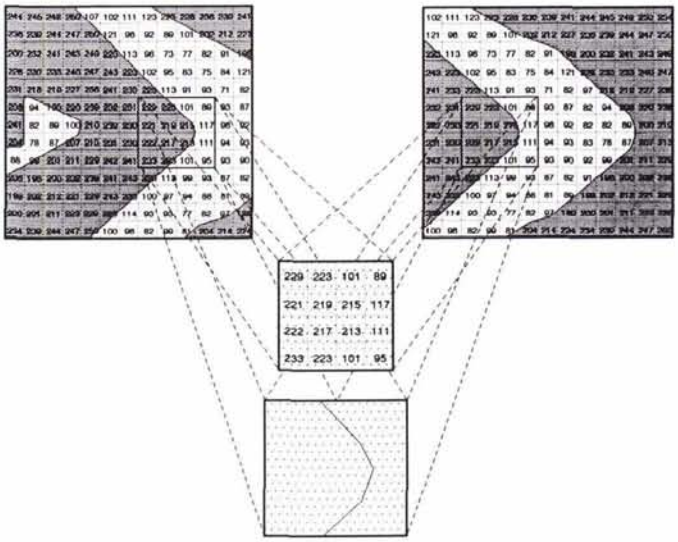

FIG. 1. Image-based matching versus feature-based matching.

the groundel attributes as well as on the availability of multiple images (to make the least squares solution meaningful) which correspond to a particular groundel. Nevertheless, the realization that the matching process takes place in the object space and not in the image space is very important.

FACTORS TO CONSIDER IN A MATCHING SYSTEM

When designing a matching system, one should decide on several key components that will determine the flexibility and the robustness of the system. These factors are

- *Matching primitive.* There are four choices for matching primitives: points, edges, areas (image patches), and semantic lists. Each primitive has its merits and its flaws. Generally, there is a direct relationship between the ability to resolve difficult matches and the complexity of the primitive used for matching. An elementary primitive such as gray level of a pixel may require imposition of various restrictions in order to ensure correct matching. These restrictions include epipolar geometry (no y parallax), close approximation of the images (x parallax of only a few pixels), close resemblance of the two images, etc. On the other hand, edges with a semantic list of their attributes can be used in a fairly unconstrained environment and still come up with the correct match.
- *Extracting primitives.* Points are principally obtained by interest operators. However, there are several computational methods to compute interest points. Some of these methods are surveyed in Luhmann and Aaltrogge (1986) and Hanna (1985). The same argument holds for extracting edges. There are numerous edge detectors that one can use. Their complexity ranges from a simple thresholding of the image to very sophisticated edge detectors such as Canny's (1986) or the LoG (Marr and Hildreth, 1980). There are also several approaches for area correlation. It can be done by computing the maximum of the correlation coefficient (Helava, 1978), or by a sophisticated least-squares solution (Helava, 1988; Ackermann, 1984; Gruen, 1985; Forstner, 1986; Wrobel, 1987). For semantic lists one needs to define the variables or the attributes which will be listed and how to obtain them (Boyer and Kak, 1986).
- *Search or prediction strategy.* The objective of this component is to reduce the search space for a match into a manageable size. Initially, each primitive in one image is a candidate for a match with all the primitives in the other image. Ideally, the search strategy should find only one candidate for a match which is also the correct one. The easiest way to reduce the search space is to impose spatial restriction on where a match can be found. Perhaps the most widely used constraint is the one which assumes compliance with the epipolar condition. Satisfying this constraint results in reducing the search space from a two-dimensional to a one-dimensional search. This is a very useful tool for reducing the search space; however, the matching system should be able to establish this condition automatically, rather than requiring a human operator to perform the relative orientation.

  In order to construct the epipolar condition automatically, one should incorporate a coarse to fine (hierarchical) search strategy in the matching system. This strategy implies that initially only a few of the more pronounced features are matched. These matches are used to establish some constraints which are imposed on the matches in the next resolution level until the original images are matched. In other words, we use intermediate matching levels to *tune in* into the correct matches by controlling the search pattern. Practical examples of how to accomplish a coarse to fine search strategy are image pyramids, low pass filtering of the images, use of different image correlators which tune into different frequency bands of the image, and approximation (simplification) of the matching primitives (Greenfeld, 1987; Hanna, 1985; Helava, 1978).

  Statistical predictions of the possible location of the next match can also be employed, especially when the matching is performed *differentially in a continuous scanning manner.* One assumes that only small changes occur between two consecutive matches; thus, based on previous matches, a new match is predicted. Matching is accomplished by verifying and refining the prediction. This technique is usually incorporated in image correlation based matching systems.

  Another means for reducing the search space was mentioned earlier and is related to the choice of matching primitives. One may expect fewer matchable complex primitives (such as continuous edges) than simple primitives (such as gray level values of a pixel). This is the main difference between SRI's stereo matching system described in Hanna (1985) and the one outlined in this paper. SRI's system uses a few good interest points as matching primitives and an intensive image pyramid strategy (eight image levels) for reducing the search space. Our system uses a few edges and only two to three image levels to accomplish stereo matching. As explained later, this is not a matter of choosing one primitive over the other, this is a conceptual difference in approaching the solution for the stereo matching problem.
- *Assessment of results.* It is rare that a primitive from one image is matched with only one primitive from the other image. This is especially true for matching systems that do not require strict similarities between the matched primitives. By allowing more *flexibility in the depiction of the features,* we enable the matching system to accommodate an extended range of existing geometrical and radiometrical distortions. Thus, the purpose of this component is to establish procedures for selecting the most likely match from the list of candidates. This is essentially a statistical process in which candidates are associated with a probability to be the correct match. Probabilities can be computed from variances (of a least-squares solutions), from weighing similarity metrics, from Bayesian analysis, and/or by computing frequency histograms of expected patterns of correct matches. The candidate with the highest probability is accepted as the correct match.
- *Quality control or consistency measures.* Matches are accepted as correct ones because, locally, they displayed the highest probability in the similarity tests. These matches, however, may not be consistent with others when analyzed in a more general inter-relationship context. Thus, the purpose of this component is to eliminate erroneous matches when examined in a global context. The use of this consistency check increases the reliability and the completeness of the solution.

It is important to realize that a matching system is not restricted to use of a single choice for each component. One may (and is encouraged to) use more than one primitive for matching or use more than one model for assessing the matching results. This will upgrade the system and make it more robust in resolving difficulties which exist in real world images. Some of these real world difficulties are presented in Table 1. The matching system should attempt to resolve all of these difficulties or at least be designed so that solutions can be realized when the will be available. These difficulties can be resolved using either algorithms or knowledge based reasoning. Algorithms can handle most of the photometric and geometric problems using well developed mathematical models for these distortions. Problems arising from the texture content of

TABLE 1. PROBLEMS ASSOCIATED WITH IMAGE MATCHING

- Photometric (radiometric) problems
  - *Resolution* due to atmospheric conditions and the camera's optics quality (especially in the corner of the image frame). The corresponding images will have different sharpness (a low pass filtering effect on one image only).
  - *Reflectance* such as sparkling of water bodies.
  - *Illumination*. Effect of the sun's angle and strength of illumination due to partial cloudiness.
  - *Foreshortened effect*. Elements smaller than pixel size which change the value of the sampled gray level (Horn, 1983).
  - *Exposure parameters differences* in the (analog) camera.
  - *Lab processing noise*. If images are digitized from the negative or a diapositive, the photographic material may have scratches or spots due to uneven processing or due to aging chemicals.
  - *Digital camera radiometric calibration* differences (integration time, gray level range definition, exposure setting, etc.).
  - *Digital camera noise* during image digitization.
- Geometric problems (Perspective projection results in:)
  - *Relief displacement* and *occluded areas*.
  - *Projective deformation*.
  - *Scale variation* due to changes in the distance between the camera and the recorded object.
  - *Base/Height ratio*. The smaller the B/H ratio, the less the effect of geometric distortions; however, the height determination is weakened and vice versa.
- Textural problems
  - Existence of *distinguishable structures*. Featureless surfaces such as ice sheets, sand, and man-made objects such as runways are extremely difficult to match (if possible at all).
  - *Repetitive texture* such as roofs, marked parking lots, plowed fields, etc.
  - *Hanging surfaces* such as multi-level highway intersections.
  - *Ambiguous levels* such as tree tops and the ground below them.
  - *Thin objects*, which are one pixel wide, may be represented differently in the pixel grid (stair case effect).

the image can be handled more effectively by tools that utilize knowledge based reasoning such as expert systems.

AN OPERATOR ORIENTED MATCHING SYSTEM

There exists one matching system that has the capabilities to overcome the matching difficulties presented in Table 1. This is the human vision system. A human operator can be trained to achieve stereo matching even under extremely involved conditions. It is only natural to infer that an automated artificial operator is a comprehensive solution to the matching problem at hand. The above is an agreeable statement; however, we must raise the inevitable question: How do we do it? How do we go about building an intelligent system that is capable of performing stereo matching automatically?

A sound starting point is to use whatever is already known about human vision and human perception, and implement it. The implementation should be done so that it will provide flexibility to future enhancement of the system as new facts become available. A fundamental fact about the human vision system is the use of edge or line queues to understand images and recognize objects. Attneave (1954) showed that a crude outline of an object provides sufficient information for the human vision system to recognize an object. Our ability to understand a cartoon in a newspaper is a proof of this theory, which we experience in everyday life. From this point of view, processing the entire array of gray levels of a digitized image during the stereo matching process seems rather redundant.

The human operator also uses a coarse to fine search strategy. Photogrammetrists know from experience that matching two images in a stereoplotter can be very difficult if there is no relative orientation and if the field of view does not contain corresponding images. A trial and error approach can prove to be time-consuming and frustrating depending on how far away are the corresponding images. To resolve such a matching problem, the photogrammetric operator will most likely follow these steps:

- Examine the two prints (not in the stereo plotter) to gain a sense of the general relationship between the two images.
- Identify, at the lowest magnification possible, some of the more distinguishable features such as buildings, intersections, or special texture forms.
- Bring the two images to close correspondence, increase magnification, and bring the floating marks to coincidence.

At first, the operator performs monocular selection of features from the image that lead to a coarse stereo matching followed by precise matching. The selection of the features is a knowledge based reasoning process that is particular for the operator and for the given stereo pair. In another stereo pair even the same operator may select other features for matching. The coarse matching is the process of bringing the corresponding features into the field of view and the precise matching is the merging of the floating marks.

Another aspect of the human stereo vision is the edge detection method. Marr and Hildreth (1980) claimed that the receptive fields of retinal ganglion cells are organized in the following manner. Light striking the center of the cell's receptive field excites the activity of the cell, while light striking the surrounding area inhibits it. The shape of the sensitivity distribution of the cells across the receptive field can be mathematically described by a Laplacian of the Gaussian (LoG). It was also found that the human vision system uses a set of *channels*, each sensitive to different frequency resolutions. Varying the size of the LoG mask (by varying the standard deviation of the LoG) imitates this characteristic of the human vision. Though experiments with the LoG edge operator show that in finer resolutions it performs less accurately than some other edge detectors such as Canny's (1986), we found the LoG to be very useful in coarser image resolutions. A very important characteristic of the LoG is its behavior in different spatial resolutions. An edge, which was detected in a coarse resolution, will not disappear as we proceed to finer ones. This continuity, or the guaranteed existence of a detected LoG edge, makes it attractive for a focusing process (Bergholm, 1987).

Finally, the human uses knowledge in resolving complex images. The stereo process itself is not an intelligent process. One can establish a stereo model of two artificially generated random dot stereograms. Thus, the human does not need any knowledge for the stereo process itself. However, in order to resolve complex ambiguity problems, such as discriminating tree tops from ground, knowledge base systems are indispensable. This is one of the most difficult and challenging task of automatic mapping using digital photogrammetry. The task is to determine what knowledge is being used, to build the knowledge base, and to develop reasoning strategies for applying this knowledge in the matching process.

AN OUTLINE OF AN OPERATOR-BASED MATCHING SYSTEM

Based on the discussions in the previous sections, we now introduce a matching system which attempts to meet the stated challenges. Figure 2 presents the block diagram of such a matcher. The ultimate goal of this matching system is to become an automated operator. While it is naive to presume that with current knowledge of artificial intelligence this goal can be realized, nevertheless, it has provisions for integrating knowledge base techniques as they become available. We limit here the discussion to the photogrammetric application of stereo matching. Expanding the system to other applications, such as remote sensing, can also be realized but are beyond the scope of this

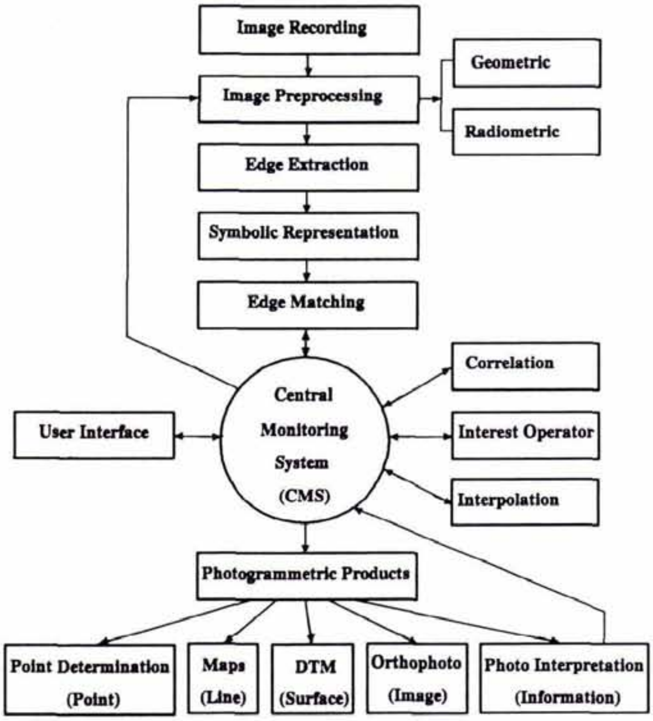

FIG. 2. An operator-based automated photogrammetric system.

paper. We now turn to explain the various components of Figure 2.

IMAGE RECORDING

Image recording is done by a sensor which records an intensity response from the scene. There are two important considerations involved in this process, that is, the sampling interval and what to record.

In selecting a sampling interval, we need to optimize two conflicting requirements. The first is that the pixels should represent a small area of the scene. The smaller the area, the more accurate the results of the photogrammetric product. On the other hand, small area per pixel requires many more pixels per image and one runs into storage and excessive computation problems. A compromise between these conflicting requirements, which is commonly used in photogrammetry, is a pixel size of 20 by 20 to 30 by 30 micrometres.

The other consideration in image recording is what to record, in other words, to record the entire image prior to any further process, or to record portions of the image as dictated by the progress of the matching. The latter is sometimes not an option but rather is compulsory, depending on the computer configuration utilized for the matching system. Storing the entire image requires an extensive storage device which may not be available on smaller computers. New technologies, such as optical disks, are likely to alleviate this storage problem. Recording and storing the entire image at this input stage has the advantage of providing more consistent pre-processing and more flexibility in implementing global measures for *tuning in* to the correct match. This consideration is, naturally, relevant only for digitizing existing analog images. If the scene is recorded directly (SPOT, Landsat, SLAR images, etc.), then the entire image is digitized.

IMAGE PRE-PROCESSING

Image pre-processing serves as an image refiner in which the raw gray level values are corrected for geometric and radiometric distortions. The geometric distortions are caused by the operating principles of the sensor, systematic errors, and distortions in the analog image (if a photograph is digitized). These distortions result in displacement of the correct location of features in the image. Distortions due to operating principles are, for example, panoramic distortions which are inherited in most space-based sensors. The sensor's systematic distortions are due to lens distortion, refraction (if the scene is digitized directly), and irregularities in the alignment of the sensoring cells in a perfect grid or line, to name only a few. For sensors utilizing a linear array, distortions may occur due to non-parallel scanning lines and jaggedness in the interval stepping of the sensor. The geometric distortions of the analog image are those which are corrected during the interior orientation in a standard photogrammetric operation. Geometric distortions are usually modeled in a calibration process and applied to the image by a transformation process.

Radiometric pre-processing is used to enhance the appearance of the image by optimizing the brightness and the contrast of the image. It is also used to remove noise from the image. The need for radiometric corrections originate from miscalibration of the digital camera (black-and-white calibration, integration time, etc.), illumination conditions, and other disturbances occurring during image recording. Common operations to improve the appearance of the images include histogram equalization for expanding the range of gray levels (which improve detectability of many image features) and low pass filtering for noise removal. Application of radiometric corrections do not alter the position of a feature on the image; however, it increases the similarity of corresponding images, and, thus, matching proficiency increases.

EDGE EXTRACTION

Edges are curves in the image where rapid changes occur in its intensity (brightness or gray level value). Physical boundaries (of objects, textures, surfaces, etc.) tend to show upon as rapid changes of the intensity in the image; therefore, we are interested in extracting edges from images. Edges are detected by edge detectors which use various mathematical techniques to mark these rapid changes in the image function.

An example of a more sophisticated edge detector is the Laplacian of a Gaussian (LoG) which we discussed earlier. This is a nondirectional operator which blurs the image (to improve the signal to noise ratio) with a Gaussian function, and marks the edge at a zero crossing of the second derivative of the blurred image. As a result of the convolution theorem, the LoG can be applied as a one-step filtering mask of the image rather than a two-step operation of blurring and differentiating. The standard deviation of the Gaussian is used to control the resolution of the edges. A large standard deviation results in coarse edges, while detailed edges are obtained with small standard deviation. The resultant convolved image is scanned to detect pixels which have a zero value or pixels at which a change of sign (positive to negative or vice versa) has occurred. These pixels are marked as edge pixels. An edge is subsequently formed by linking adjacent edge pixels into a continuous edge contour.

SYMBOLIC REPRESENTATION

An edge contour contains an enormous number of individual edge pixels. It is not practical and not necessary to use all of these pixels for subsequent processes, such as finding corresponding edges. Instead, it is necessary to represent the edge with a more simple form such as a string of nodes (polygon vertices) of characteristic edge pixels. Thus, rather than matching the curved edges, we only have to match these polygons. Another aspect of symbolic representation is to establish a list of attributes to characterize a specific edge. This list is very useful in the matching process as matches will be carried out only between edges that have similar characteristics. Greenfeld (1989) provides an elaborate discussion on symbolic representation.

Edge Matching

Up to this stage all operations have been performed on each image separately. Information from each image was gathered and compiled into a meaningful form which will assist us in building a stereo model. Edge matching is the crucial step in which corresponding edges (or their symbolic representation) of the stereo pair are matched to establish a "three-dimensional wire frame" of the scene. Edge matching is performed by analyzing similarities between each edge on one image and those from the other one. In addition to similarity measure, one should employ consistency and continuity measures as well, to ensure that a local similarity between two features is in agreement globally with other matches. The similarity, consistency, and continuity measures have been discussed earlier.

Central Monitoring System (CMS)

This is the principal component or the heart of the matching system. It operates similarly to an inference engine of an expert system. Its responsibility is to coordinate and supervise the entire photogrammetric process. In other words, the goals and functions of the CMS are to interpret the user's request, assess the task at hand, handle efficiently specific operations such as feature matching, resolve difficult matching problems, and produce the requested photogrammetric product. Consequently, the CMS is used to assign tasks to different modules of the system and to provide these modules with the proper data and information necessary for their operation. In turn, the various modules of the photogrammetric system provide information and feedback to the CMS in order to verify findings and in order to be assisted in solving problems arising during task performance. Verification of findings may be needed, for example, in object recognition and photo interpretation or in edge matching using object hypotheses (matching two sets of edges that represent the same object). An example for assistance needed in solving an operational problem is the case of loosing "stereo lock" during data acquisition for a digital terrain matrix (DTM) by image correlation. The CMS may be requested to restore stereo lock from edge (or other) information. Building this CMS is an enormous challenge in digital photogrammetry. It will take several years until it will be fully realized.

User Interface

The user interface has two functions. First, it may be used to input data needed for interior and absolute orientation. The interior orientation data are needed for the preprocessing stage to correct geometrical distortions of the image. The absolute orientation data are not essential for the data compilation process. Data can be collected in a local model coordinate system, and the transformation to this object space coordinate system can be carried out just before delivery of the final product.

The second function of the user interface is to define for the system the nature of the required photogrammetric product and its specifications. Essentially, this is the same information provided to a human operator as a job definition. The CMS may interact with the user by requesting additional information in order to clarify some vague or ambiguous instructions.

Correlation

Generating a DTM by means of correlation methods is well documented in photogrammetric literature. This method has been used, in one way or another, for over three decades for image matching in photogrammetry. In our system correlation is performed subsequently to edge matching so that edges can be used to break up images into smaller regions and, thus, constrain the correlation process. The CMS feeds into the correlator the corresponding image patches.

Interest Points

In this solution, distinct points are extracted from two closely matched windows. The points are matched by analyzing their distinctness, similarity, and consistency with other matches. This module is invoked by the CMS, if point positioning is required. As in the case of correlation, image patches which are already closely matched are supplied to the interest operator module by the CMS.

Interpolation

This module complements the correlation module in generating a dense DTM. In the event of poor texture content in the corresponding images, correlation methods will fail to perform the matching. To obtain height values for points in this case, a mathematical surface interpolation can be employed instead of correlation. The surface is computed from the three-dimensional information of the matched edges which surround these image patches and interest points.

In summary, the above described matching system is founded on a core program which we called the central monitoring system (CMS). The input to this program includes a list of matched edges and other relevant information required to define the nature of the desired photogrammetric product. The matched edges serve as a guide for selecting corresponding image patches which are approximately matched for specific operations such as precise pointing or correlation. Some edges are also the planimetric features of a map (though a study on their spatial accuracy needs to done), so that with a sophisticated image interpretation system it may become an automatic mapper.

EXPERIMENTS

The system as presented above cannot yet be fully realized. The reason is lack of knowledge about the cognitive and intelligent process that take place in the mind of the human operator. The inability to realize the full system does not have to discourage us from implementing those tasks that can be realized. These tasks, however, ought to be consistent with the outline described in the previous sections. An example for such a task is to provide automatically closely matched image patches for interest point matching or DTM generation by correlation. Another task is to establish automatically a relative orientation between overlapping images. These tasks are currently performed semi-automatically with some human operator assistance. Elimination of the need to be dependent on the operator to perform these tasks can, for example, completely automate the aerial triangulation process.

We selected the task of automatically establishing the relative orientation as a case study. During this process, the system establishes automatically an approximate match between the images and then makes precise pointing with an interest operator. No human operator intervention is required, not even a single initial approximate stereo fusion. It should be noted that current matching systems assume that the images are already relatively oriented, or that say 70 percent of the image patches contain corresponding images (Forstner, 1986).

Table 2 summarizes our choice of means to achieve our goal. The elements of the figure follow the discussion presented in the third section. The main point to notice in the figure is that the selection of means is consistent with what is believed to be used by the human operator as discussed in the fourth section.

Figure 3 describes the subset of Figure 2 which was realized in the implementation of our matching system. Accordingly, edges are matched and the central monitoring system extracts corresponding image patches from the stereo pair. The image

TABLE 2. COMPONENTS OF THE SYSTEM TO ESTABLISH AUTOMATICALLY THE RELATIVE ORIENTATION.

Component	Implementation.
Primitives for matching:	Edges and points.
Primitive extraction:	For edges LoG, for points Forstner's (Paderes et al., 1984) interest operator.
Search strategy:	Three coarse to fine techniques. • Image pyramids. • Different spatial resolution of the LoG. • Different resolution of edge approximation. Use of semantic attribute list (orientation, angles, edge strength, and relative location).
Assessment of matching results:	Probability analysis of matching candidates. Probabilities obtained from frequency histograms of spatial relationship among the matching candidates. For details see Greenfeld (1989).
Quality control:	Residual analysis of remaining y parallaxes after relative orientation adjustment.

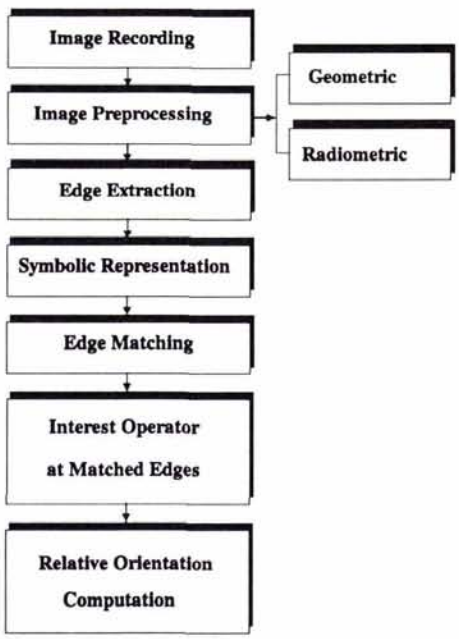

FIG. 3. Implementation aspects of the system to establish automatically the relative orientation.

patches are then processed in the interest operator program. The image coordinates of the matched points are used as observations in a least-squares solution for the relative orientation. More details about the exact implementation of the system can be found in Greenfeld (1987, 1988).

RESULTS

Results of the relative orientation are presented in Table 3. The first row in the table presents the initial average y parallax of the model in the zero assumption, i.e., applying no *a priori* knowledge about the relative alignment of the two images. The second row (comparator) presents the orientation elements obtained from points measured by a human operator on a stereocomparator. These values serve as ground truth for the automatic procedure. The fourth row (points) presents results as obtained automatically by our matching system. It is evident from the results that the orientation, which was determined completely without human intervention (digitized with approximately 0.1-mm pixel size), is comparable to the one obtained by the human operator. The result of the automatically computed relative orientation is based on 78 matched points. Thus, the statistics may be considered as highly reliable (73 degrees of freedom) as only five matches are required for solving the relative orientation problem. The orientation elements are in degrees of arc, while Py (the residual y-parallax in the model) is in pixel units.

TABLE 3. RELATIVE ORIENTATION RESULTS (ORIENTATION ANGLES IN DEGREES AND PARALLAX IN PIXEL UNITS).

	ω'	φ'	κ'	ω''	φ''	κ''	Py	σ_{Py}
Initial	0.0	0.0	0.0	0.0	0.0	0.0	123	0.0
Comparator	0.0	−0.03	−0.52	4.37	0.20	−1.19		
Vertices	0.0	−1.67	−0.68	3.96	0.37	−1.17	6	
Points	0.0	−0.03	−0.43	4.41	0.23	−1.12	0.10	0.35

Figure 4 presents the image for which the relative orientation has been computed. The inexact correspondence between the detected edges and the true edges is due to the spatial extent of the LoG. The displayed edges were obtained using a 128 by 128 LoG mask which makes up our coarsest edge resolution. For the purpose of relative orientation, it proved to provide a sufficient number of edges to enable the interest operator to work well. The advantage of such a coarse edge map is that we don't have to deal with many edges in the search for matches.

CONCLUSIONS

A conceptual model for a comprehensive automated photogrammetric system has been presented. The matching system is centered around an artificial photogrammetric operator. Thus, the system is based on the idea of mimicking the best known working matching system, namely, the human photogrammetric operator. It attempts to incorporate existing knowledge on how the human operator performs stereo matching and existing knowledge on relevant methods for image matching. It is not a system for merely finding similar images but for doing it in a context of a larger automated photogrammetric system.

The implication of an operator oriented approach is that processes, primitives, and strategies used by the human operator have to be implemented into the matching system. For example, the use of edges as the leading matching tool is one consequence of using the human operator as a model for the matching system. Another example is the use of coarse to fine strategy as a *tuning in* device, from familiarization with the content of stereo pair to precise point positioning. The ultimate goal is to provide the system with a job description and other relevant data, similar to the one provided to a human operator, and the system will perform the job.

A simplified version (without the intelligent component) was implemented to perform automatically the relative orientation. The objective was to perform the relative orientation without any intervention of a human operator at any stage. Thus, the system receives as input two digitized aerial photographs, data about interior orientation constants, and a rough estimate of the overlap percentage. Experiments have been made and results have been presented to support the feasibility of accomplishing a fully automatic relative orientation based on this minimal information.

The implementation of the intelligent component will continue to challenge the photogrammetric, artificial intelligence, and computer vision communities for years to come. It will require, among others, that expert systems be built for specific tasks and that they be integrated later into a comprehensive intelligent system.

REFERENCES

Attneave, F., 1954. Some Information Aspects of Visual Perception, *Psychological Review*, 61: 183–193.

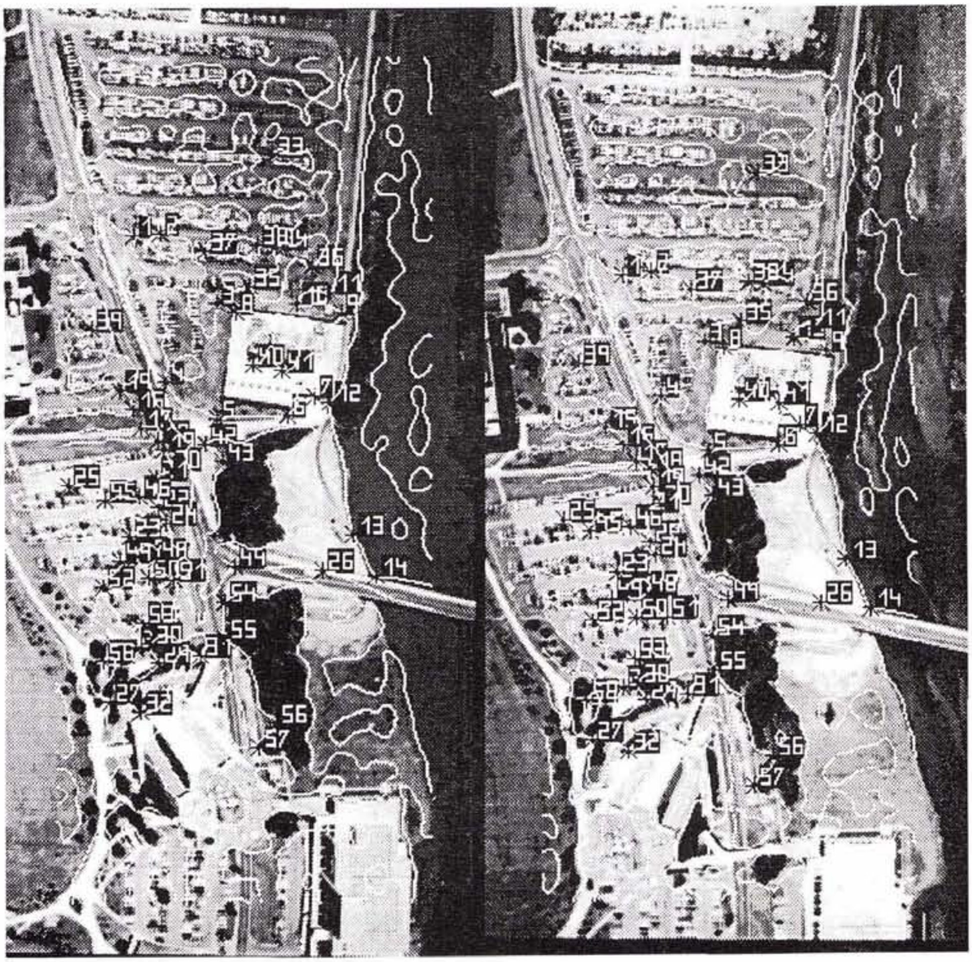

FIG. 4. Example of matches obtained during the relative orientation process.

Akermann, F., 1984. Digital Image Correlation: Performance and Potential Application in Photogrammetry, *Photogrammetric Record*, 11(64), 429–439.

Bergholm, F., 1987. Edge Focusing, *IEEE PAMI-9*, No. 6, 726–741.

Boyer, K. L., and A. V. Kak, 1986. *Symbolic Stereo from Structural Descriptions*, Report TR-EE-86-12 Purdue University, W. Lafayette, Indiana.

Canny, J. F., 1986. A Computational Approach to Edge Detection, *IEEE PAMI-8*, No. 6, 679–698.

Forstner, W, 1986. A Feature Based Correspondence Algorithm for Image Matching, *Int. Archives for Photogrammetry and Remote Sensing*, Vol. 26–3/3, 150–166.

Greenfeld, J. S., 1987. *A Stereo Vision Approach to Automatic Stereo Matching in Photogrammetry*. Report No. 381, Department of Geodetic Science and Surveying, The Ohio State University, Columbus, Ohio. Also, PhD dissertation, Department of Geodetic Science and Surveying, The Ohio State University, Columbus, Ohio.

———, 1988. Stereo Matching Without an Operator's Intervention, *Proceedings of 1988 ACSM-ASPRS Annual Convention*, St. Louis, Missouri **Vol. 3**, pp. 56–65.

———, 1989. Experiments with Edge-Based Stereo Matching, *Photogrammetric Engineering & Remote Sensing*, Vol. 55, No. 12, pp. 1771–1777.

Grimson, W. E. L., 1981. Theory of Human Stereo Vision, *Phil. Trans. Royal Society of London*, B 292, pp. 217–253.

Gruen, A, 1985. Adaptive Least Squares Correlation: A Powerful Image Matching Technique, *South Africa Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3), 175–187.

Hanna, M.J., 1985. SRI's Baseline Stereo System, *Proceedings of the DARPA Image Understanding Workshop*, Miami Beach, Florida, pp. 149–155.

Helava, U. V., 1978. Digital Correlation in Photogrammetric Instruments, *Photogrammetria*, Vol. 34, pp. 19–41.

———, 1988. Object Space Least Squares Correlation, *Proceedings of 1988 ACSM-ASPRS Annual Convention*, St. Louis, Missouri, **Vol 3**, pp. 46–55.

Horn, B. K. P., 1983. Non-correlation Methods for Stereo Matching, *Photogrammetric Engineering & Remote Sensing*, Vol 49, No. 4, pp. 535–536.

Julesz, B., 1960. Binocular Depth Perception of Computer Generated Pattern, *Bell Syst. Tech. J.*, Vol. 39, pp. 1125–1162.

Luhmann, T., and G. Aaltrogge 1986. Interest Operators for Image Matching, *Proceedings of ISPRS Comm. III*, Rovaniemi, Finland, pp. 459–474.

Marr, D., 1982. *Vision*, W. H. Freeman and Company, San Francisco, California.

Marr, D., and E. C. Hildreth 1980. A Theory of Edge Detection, *Phil. Trans. Royal Society of London*, B 207, pp. 187–217.

Paderes, P. C., Jr., E. M. Mikhail, and W. Forstner, 1984. Rectification of Single and Multiple Frames of Satellite Scanned Imagery Using Points and Edges as Control, *NASA Symp. on Math. Pattern Recogntion and Image Analysis*.

Wrobel, B., 1987. Facet Stereo Vision (FAST Vision)-A New Approach to Compute Stereo Vision and to Digital Photogrammetry, *Proceedings of Intercommission Conference on Fast Processing of Photogrammetric Data*, Interlaken, pp. 231–258.

(Received 10 July 1989; revised and accepted 17 September 1990)