Image Matching Based on Tracking Matching Paths in the Similarity Space

Xueliang Mei, Peng Gong, and Greg S. Biging

Abstract

In this paper, we explore the feasibility of implementing surface continuity and ordering constraints by tracking connected local peak points in the space of matching similarity. We present a new image matching algorithm based on match paths directly extracted in the similarity space. The main advantage of these kinds of tokens over tokens extracted from the left and right images is that the geometric distortions caused by perspective effects and description inconsistency caused by independent extraction of tokens in the left and right images are automatically eliminated. Through the tracking process, the computation is greatly reduced. In addition, global information available to support a local match for resolving matching ambiguities is fully utilized in such a way that unrelated global information is excluded. Thus, the new image matching algorithm is reliable and efficient. By eliminating the interpolation process at the levels except the finest, the occluded regions and depth discontinuities can be well localized.

Introduction

One of the vital problems in automatic DEM (digital elevation model) generation is to match a set of identifiable physical points over a stereo pair to derive a disparity map. This problem is known as the image matching problem and the physical points are referred to as matching tokens (or features). Image matching is an integral part of automatic DEM generation in digital photogrammetry. Many solutions have been proposed (Barnard and Fischler, 1982; Dhond and Aggarwal, 1989; Heipke,1996).

A general image matching procedure includes four steps:

- Extract matching tokens (e.g., intensity windows, edges, and linear segments) and their descriptions;
- Compute similarity values based on the descriptive information of a token to find candidate matches in a search range;
- Resolve matching ambiguity by employing constraints such as surface continuity and ordering constraints (Jones, 1997) through a global process; and
- Interpolate unmatched points to derive a final dense disparity map.

It is not an easy task to develop a successful image matching algorithm. Many factors should be taken into consideration. The diversity of matching tokens is large, and they should be carefully chosen. Common matching tokens include intensity windows (Grün, 1985; Rosenholm, 1987; Kang *et al.*, 1994; Luo and Burkhardt, 1995), groundels (Helava, 1988), edges (Grimson, 1985; Hongo *et al.*, 1996), linear segments (Medioni and Nevatia, 1985; Ayache and Faverjon, 1987; Horaud and Skordas, 1989; Sherman and Peleg, 1990), and relational structures (Shapiro and Haralick, 1987; Christmas *et al.*, 1995). The choice depends on the image data and the characteristics of the imaged physical scene. For example, in urban areas, geometric structures are rich and, thus, edges and linear segments are more suitable matching tokens than are other kinds of matching tokens while, in open areas, textures are rich and, thus, intensity windows are a good choice. Because similar tokens occur frequently over a token's search range, matching ambiguity should be resolved by employing constraints. Identified constraints include the epipolar constraint (Baker and Binford, 1981; Jones, 1997), the surface continuity constraint (Grimson, 1981; Jones, 1997), the figural constraint (Jones, 1997), the ordering constraint (Ohta and Kanade, 1985; Jones, 1997), and the hierarchic structure constraint (Shapiro and Haralick, 1987; Jones, 1997). Among these constraints, the most difficult constraint which can be augmented into an algorithm is the continuity constraint because in the real world this constraint is frequently violated by depth or orientation discontinuities. Methods such as relaxation (Christmas *et al.*, 1995) or artificial neural networks (Cruz *et al.*, 1995; Grant *et al.*, 1998) can be employed to implement the continuity constraint through a compatible relationship description model or a weight connection description model derived from the continuity and other constraints among the potential matches. In these methods, the local information is propagated to other places through iteration. In order to increase the reliability and efficiency of the image matching process, a hierarchical strategy such as from coarse to fine (Terzopoulos, 1983; Rohaly and Wilson, 1993) or multi-primitives (Marapane and Mohan, 1994; Venkateswar and Chellappa, 1995) is frequently applied. In this mechanism, the matches from the coarser levels are used to predicate matches at the finer levels. Thus, the search range is decreased and image matching becomes more reliable. Correspondingly, the computation is also greatly reduced. The main problem in this mechanism is that it is difficult to control error propagation from the coarser levels to avoid blunders in the final disparity map. A combination of interpolation and matching in grids having some intervals or employing high level tokens such as edges, linear segments, and relational

Center for Assessment and Monitoring of Forest and Environmental Resources, Department of Environmental Science, Policy and Management, 151 Hilgard Hall, University of California, Berkeley, CA 94720-3110 (gong@nature.berkeley.edu).

X. Mei is currently with Michael Baker Corporation, 4301 Dutch Ridge Road, Beaver, PA 15009 (xuelm@yahoo.com).

Photogrammetric Engineering & Remote Sensing
Vol. 67, No. 4, April 2001, pp. 453–460.

0099-1112/01/6704–453$3.00/0

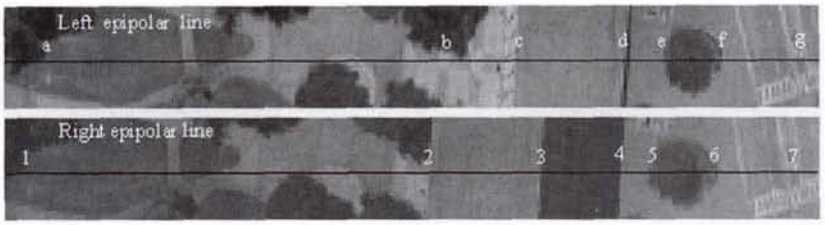

Figure 1. An epipolar stereo pair (the black lines are the corresponding epipolar lines used to generate Figure 2).

structures instead of matching every pixel can further reduce the computation and improve matching reliability. Other factors include matching uncertainties caused by perspective distortions between the left token and its corresponding right token and by signal noise during the imaging and scanning process. The solutions to this problem may be referred to as the shaping window technique (Okutomi and Kanade, 1992).

In this paper, we implement the surface continuity and ordering constraints by tracking connected local peak points in the matching similarity space in a new image matching algorithm. The algorithm is a "from-coarse-to-fine" algorithm based on match paths directly extracted by tracking connected local peak points in the similarity space. Our algorithm was implemented on a low-end DEC 3000 Alpha workstation. Two stereo pairs—an urban scene and an oak woodland scene—were used.

Image Matching by Tracking Connected Local Peak Points in the Similarity Space

The surface continuity constraint rests on the observation that the real world can always be considered as piecewise continuously smooth. In practice, this may be formulated in a number of different ways. For example, two neighboring points in the image space should have a "similar" disparity, i.e., the disparity difference should be small and within some threshold. We can construct an image in which the *I*-axis represents an epipolar line in the left image and the *J*-axis represents its corresponding epipolar line in the right image. A pixel (i, j) in this space represents a potential match. Every pixel is filled with a similarity value. The brighter the pixel, the more likely that the pixel is a correct match. We may find that bright pixels, especially those pixels at local peak points along their columns, will be clustered together to form bright segment paths in the similarity image because of the surface continuity constraint. As an example, Figure 1 shows that segments of **ab/12**, **cd/23**, **de/45**, and **fg/67** lie on continuously smooth surfaces. Figure 2 is a similarity space constructed based on the two black epipolar lines shown in Figure 1. The brightness values are cross-correlation coefficients calculated from the epipolar lines. In Figure 2, the cross-correlation coefficient image presents salient bright paths of **AB**, **CD**, **EF**, and **GH**. Those paths appear to be connected local peak points in the column direction. The correspondence between segments that lie on continuously smooth surfaces and those segment paths in the similarity space in the form of connected local peak points makes it possible to treat image matching as a bright-path tracking problem. We can benefit from turning image matching into a bright-path tracking problem in three aspects:

- We do not need to employ computationally complex shaping window techniques (Okutomi and Kanade, 1992) or adaptive least-squares image matching (Grün, 1985) to address the matching uncertainty problem caused by the geometric and radiometric distortions between a left token and its corresponding right token. In the tracking process, the geometric distortions are implicitly eliminated, and the extracted segment paths in the similarity space correspond automatically to two segments having geometric distortions in the left and right image. For example, segment **AB** in Figure 2 corresponds to segments **ab** and **12** in Figure 1. Segment **ab** is a little shorter than segment **12**.
- Traditionally, features such as edges (Grimson, 1985; Hongo *et al.*, 1996) or linear segments (Medioni and Nevatia, 1985; Horaud and Skordas, 1989; Sherman and Peleg, 1990) are first extracted from the stereo pair, then image matching based on these features is performed. This preprocess makes image matching error-prone because the feature extraction process cannot guarantee extraction consistency between the left and right images. For example, in the left image a physical linear segment may be extracted partially and in the right image this physical linear segment may be over-extracted or split into several small segments. The inconsistency between the left features and their corresponding right features makes these features unmatchable. Features such as segment paths that are formed by local peak points in the similarity space always correspond to those segments physically lying on the continuously smooth surfaces in the left and right images. Because these features are extracted once, the consistency between the corresponding left and right features is always preserved.
- Global matching evidence can be fully collected while unrelated global matching evidence is ignored, leading to higher matching reliability. The tracking process will stop at places where

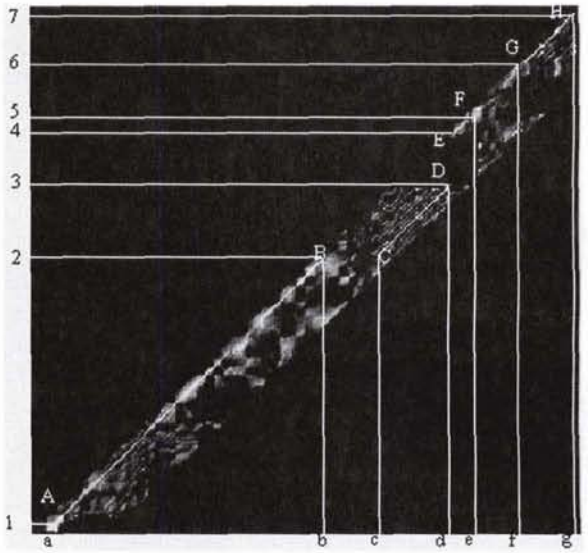

Figure 2. The correspondence between segment paths formed by local peaks along the column direction and the connected matching paths in the similarity image.

occlusions or different continuously smooth surfaces occur. Thus, the segment path extracted by such techniques will always lie on the same continuously smooth surface. Because different continuously smooth surfaces are unrelated, a correct match on a continuously smooth surface will not provide supportive matching evidence to another match on a different surface. In Figure 1, segment **ab** and segment **cd** are not related; segment **ab** is on a polynomial ground surface while segment **cd** is on a flat building top.

The ordering constraint is based on the observation that the order of any two points (i.e., left to right and right to left) in an epipolar plane in the real world will remain in the stereo projection. This makes the tracking more feasible because correct segment paths will extend along the diagonal. Vertical or horizontal segment paths are those paths which correspond to occlusions or "NULL" matches.

Tracking is not feasible under some circumstances. In Figure 3, if the minimum distance between any two neighboring local maximum points along the column direction of the similarity image is too small, then segment paths that are formed by connected local peak points will be clustered into bright regions of non-separable segments. This phenomenon occurs in texture-free images. Integration of shape-from-shading (Pankanti and Jain, 1995) may be a possible solution. Another case is that the corresponding matching tokens between the left and right images are not sufficiently similar. Thus, no bright segments can be formed in the similarity space. This phenomenon occurs when the perspective distortions between the left and right images are very large. For example, tokens on a vertical building wall or a deep slope of a tree crown will not appear similar in the left and right images. Thus, they are dark points in the similarity space. Employing multi-baseline stereo images (Fua, 1997) may help solve this problem.

The characteristics of the similarity image not only depend on the characteristics of the imaged scenes but also on the definition of the similarity measure. Two typical scenes can be found in the real world, natural scenes and urban scenes. In our experiments, stereo pairs of an urban scene and a natural scene were tested. There are many definitions of similarity measures. We tested three similarity measures. They are the absolute difference of gray value, the absolute difference of average gray value, and the cross-correlation coefficient. We found that the distance between any two local peak points along a column in the similarity space of the cross-correlation coefficient is the largest, and that most formed bright paths are separable. Therefore, we adopted the cross-correlation coefficient as the similarity measure for this study.

The New Image Matching Algorithm

Before the algorithm is run, a stereo pair of the epipolar image pyramids must be prepared in advance. This is done through image resampling. At each level, a line-by-line based matching is executed. First, the algorithm initializes the corresponding disparity map at that level. Second, the whole image pair at that level is processed line by line. The final disparity map is obtained through a filtering process along the y-direction. The disparity profile along each epipolar line is derived in three steps. At the first step, a similarity image is constructed. At the second step, paths formed by connected bright spots in the similarity image are tracked. At the third step, the tracked paths are used to derive the final matches whose brightness values are the highest. The final matches must also satisfy the ordering and smoothness constraints. The new algorithm is a "from-coarse-to-fine" algorithm (Figure 4). It is developed to recover scenes with weak depth discontinuities, especially scenes of wild lands. For scenes with some strong depth discontinuities, for example, urban scenes, an interpolation process at coarser levels except the finest level should be excluded. The reason for not using the interpolation process at the coarser levels for scenes with strong depth discontinuities is that unmatchable regions often correspond to the occluded regions. Their disparity values at the coarser level may mislead image matching at the finer level. The idea here is that points on continuous surfaces should be matched while the occluded points should be ignored during the image matching process.

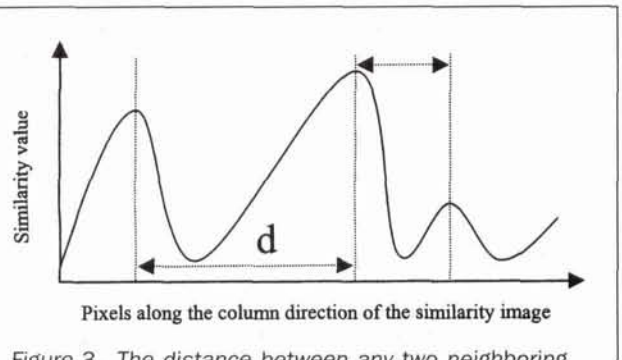

Figure 3. The distance between any two neighboring local peak points along a column.

The Number of Hierarchical Levels

The number of hierarchical levels depends on the search range, the computation, and the control of error propagation. Theoretically, 3- by 3-pixel-window averaging requires the lowest computation. However, a small averaging window causes too many levels. The more the levels of the hierarchical image pyramid, the more difficult it is to prevent errors at the coarser levels from propagating to the finer levels. Thus, very small averaging window sizes should be avoided. In our algorithm the averaging window size is 5 by 5 pixels. Thus, the disparity search range (predicated by the coarser level) at each level will be within 5 pixels, and its corresponding global maximum-minimum disparity search range is computed with the following formula:

$$D_{\max}/5^I \geq D \geq D_{\min}/5^I. \tag{1}$$

Here $D_{\max}$ and $D_{\min}$ are the global maximum and minimum disparity values at the finest level, $I = 1, 2, \ldots, N$; $N + 1$ is the number of hierarchical levels; it is computed using

$$N = \log_5^{(D\max - D\min)}. \tag{2}$$

The global maximum-minimum disparity search range at each level is used to bound the tracking process into a diagonal strip in the similarity image.

Tracking Seeds

Tracking seeds can be those points that are local peaks in the similarity image along the column direction because local peaks are most likely to be correct matches. In order to reduce the number of "bad" local peaks, a conservative threshold is used. In each column of the similarity image, only one pixel whose value is above the threshold and has the highest value in the disparity search range of this column is considered as a tracking seed. Some columns will not have tracking seeds if no local peaks exceed the threshold value. The advantage of this process lies in the fact that tracking seeds will not be clustered into regions and be limited by the number. At the coarsest level, the disparity search range is computed by the global maximum

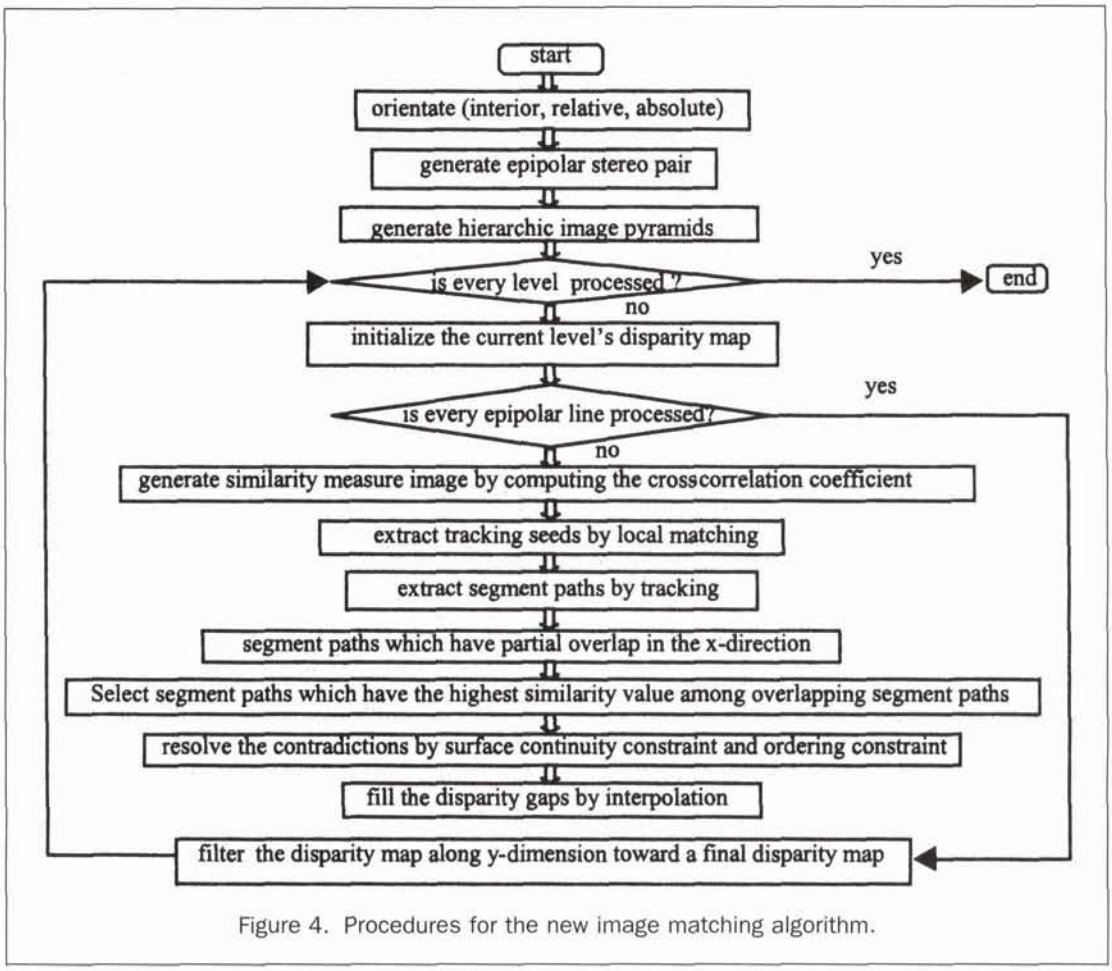

Figure 4. Procedures for the new image matching algorithm.

and minimum disparity values, while, at the finer level, the disparity search range is predicted by the disparity map of the precedent matched coarse levels and is within 5 pixels.

Although it is still possible for this process to throw away potentially correct matches and include error matches, only a small percentage of those seeds could be wrong matches, especially at the coarsest level. The following process of tracking can recover those omitted potentially correct matches by propagating other correct matches in the same continuously connected path to those points.

Tracking Process

A scan is first made from the left to the right and from the top to the bottom in the similarity image. If a seed is found, then in its compatible neighboring region a local peak is labeled as the next tracking point. This procedure continues until no peak in the neighboring region exists. The tracking process is bi-directional, i.e., the path segment before the seed and that after the seed are both tracked. During the tracking process, if other seeds are met, these seeds are labeled as processed seeds to avoid repetitive tracking; if the current path meets other paths that are already tracked out, then the tracking process stops to claim an intersection of two paths. The compatible neighboring region is defined by the following equations:

$$\|(x_1 - x_0)\| < T_1 \text{ and } \|(y_1 - y_0)\| < T_2 \quad (3)$$

$$x_1 > x_0 \text{ and } y_1 > y_0 \quad (4)$$

$$x_1 < x_0 \text{ and } y_1 < y_0 \quad (5)$$

Here (x_0, y_0) is the current point, (x_1, y_1) is the next tracking point, and T_1 and T_2 are thresholds. In our experiments, $T_1 = 1$ and $T_2 = 2$. Equation 3 is a representation of the surface continuity constraint, while Equations 4 and 5 are representations of the ordering constraint. Because the ordering constraint is affected by the image resolution, i.e., the coordinates are integer numbers while not continuously real numbers, Equations 4 and 5 should be modified into the following equations including an equality:

$$x_1 > x_0 \text{ and } y_1 \geq y_0 \quad (6)$$

$$x_1 < x_0 \text{ and } y_1 \leq y_0 \quad (7)$$

In order to avoid occurrence of long vertical segments which represent occlusions in the left epipolar line or horizontal segments which represent occlusions in the right epipolar line, the times of continuous occurrence of "equality" should be limited.

The next tracking point is the point whose value is above a threshold and has the highest value in the region defined by Equations 3, 6, and 7.

Path Dividing and Sorting

In order to simplify our tracking process, two paths sharing common parts of x coordinates are divided into four short paths. The dividing is a recursive process because some paths have several component paths that share parts of the x coordinates, especially at the finest level. At the coarser level, this process is occasionally needed. Through this dividing process,

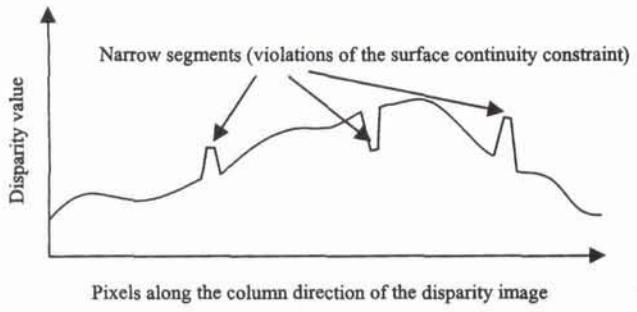

Figure 5. A disparity profile along the column direction of the disparity map.

paths can be arranged according to their castings along the x-axis. Paths which share the same x coordinates are candidate path matches.

The Similarity Value of a Path

The similarity value along a path is a sum of similarity values of its connected points. In our experiments, the similarity measure is chosen to be the cross-correlation coefficient. Thus, the similarity value along a path is a sum of cross-correlation coefficient values of its connected points.

Selection of Paths

Methods such as relaxation (Christmas *et. al.*, 1995), artificial neural networks (Cruz *et. al.*, 1995; Grant *et. al.*, 1998), and dynamic programming (Ohta and Kanade, 1985) can be used to resolve matching ambiguities among candidate paths. In fact, these computations are not necessary because the tracking process already collected global matching evidence available for a local point. Here we use a heuristic method to resolve matching ambiguities among candidate paths. The correct match among several candidate paths is selected by finding the path with the highest similarity value. In order to remove some poor matches, we use an average similarity threshold to detect bad matches. The average similarity value is computed by dividing the sum similarity value by the number of points forming that path. Then violations of continuity and ordering constraints are detected. The violation of the continuity constraint is detected by an interpolation method. If there is a large distance between the interpolated path and that to be checked and the distance is beyond the range defined by Equations 3, 6, and 7, then the path is replaced by the interpolated path. For two paths that violate the ordering constraint, the longer path is preserved and the shorter one is discarded. The unmatched gaps are filled by interpolation. Our experiments show that this kind of processing not only leads to competitive results compared with those algorithms such as global optimization but also greatly reduces the computation.

The Filtering Operation

Because we use paths as our matching tokens, errors will be segments along the x direction in the final disparity map. These errors lie in the fact that our algorithm is an epipolar-line-based method, and the surface continuity constraint between two neighboring epipolar lines is not implemented. Optionally, we may extend the current algorithm into a region-based algorithm while regions are extracted from the 3D similarity space. But, as we can see, this solution needs lots of memory practically, it is not feasible. Fortunately, these violations can be easily detected along the y direction of the disparity map through filtering. In a disparity profile along the y-direction, the narrow segments which have very different disparity values from those wide segments immediately adjacent to them are violations of the surface continuity constraint (Figure 5). Thus, they should be replaced by interpolation using wide segments immediately adjacent to them. In our filtering operation, every profile along every column of the disparity map is first classified into narrow segments and wide segments. Then those narrow segments violating the surface continuity constraint are replaced by interpolation. In our experiments, the narrow segments are within 5 pixels and wide segments are those longer than 5 pixels.

Testing the Algorithm

In this section, the efficiency and the reliability of our algorithm are examined. Two stereo pairs are used for this purpose (Figures 6 and 7). Figure 6 is a stereo pair patch of an oak woodland with a scale of 1:12,000 and at a 25-μm pixel resolution, while Figure 7 is a stereo pair patch of an urban area with a scale of 1:2400 at a 125-μm pixel resolution. Because the scale of Figure 7 is very large, occlusion phenomena are very heavy and the disparity map is difficult to derive from previous algorithms. The disparity range is $+20$ to -40 pixels for Figure 6 and $+10$ to -40 pixels for Figure 7. In Figure 6, 92 percent of the pixels have a "similar disparity within a ± 1-pixel difference" of their immediate neighboring pixels that physically lie on the same continuously smooth surface. In Figure 7, 56 percent of the pixels have a "similar disparity within a ± 1-pixel difference," and 35 percent have a "similar disparity between 1 and 2 pixels absolute difference" of their immediate neighboring pixels that physically lie on the same continuously smooth surfaces. The statistics shows that the scene in Figure 6 is smoother than that in Figure 7.

Our algorithm is programmed using C language on a DEC 3000 workstation. Figures 6C, 7C, and 7D are the derived disparity maps. From these disparity maps, we can easily find that our algorithm can work in a wide array of complex images. Continuously smooth surfaces are successfully recovered. Depth

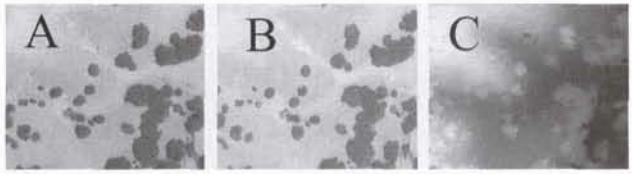

Figure 6. A stereo pair of oak wood land and its disparity map. A: left image. B: right image. C: the disparity map referenced to the left image.

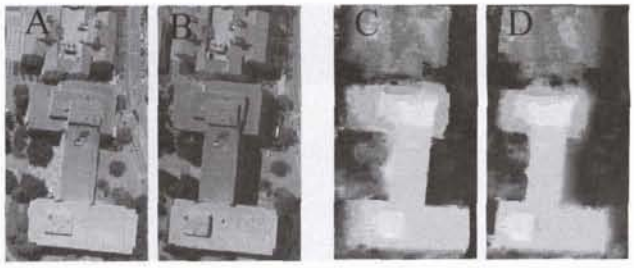

Figure 7. A large-scale urban area stereo pair. A: left image. B: right image. C: the disparity map referenced to the left image. D: the disparity map referenced to the right image.

discontinuities and occlusions are well localized and preserved in Figure 7C and 7D. These features are essential for automatic mapping of urban areas (Weidner and Förstner, 1995). Most algorithms in commercial softcopy photogrammetry systems lack the ability to deal with depth discontinuities and occlusions because, in their algorithms, grids of some constant intervals are used as matching tokens, and points that are not grid points are not matched. In the final disparity map, disparity values of those points that are not matched are interpolated from matched grid points, while the interpolation process is a "blind" process because of the lack of knowledge of the surface characteristics, i.e., irrelevant matched grids that physically lie on different surfaces may be improperly used to interpolate an unmatched point. This occurs where depth discontinuities and occlusions exist. As a result, salient important features are smoothed. In our algorithm, the tracking will stop when occlusions and depth discontinuities occur because occlusions have "NULL" matches and they are dark points in the similarity space. Tree stands in Figure 6 are well recovered but not in Figure 7. Tree stands in Figure 6 are oaks with dense leaves and the crown surfaces are relatively flat, while tree stands in Figure 7 are heavily oblique, irregular, and semitransparent. The geometric distortions between a left token and its right token that lie on the tree crowns cannot be easily modeled. Additionally, the matching difficulty is partially attributed to the large scale of the stereo pair. We sampled several epipolar lines which pass through tree crowns in Figure 7 and found segments that represent tree crowns had no corresponding salient bright segment paths in the similarity space. Only some bright points that cannot be connected scatter occasionally in those sections where tree crowns occur. This may be alleviated by employing multi-baseline stereo images. As pointed out earlier, under this special case our algorithm does not work well. As a matter of fact, this is a case that cannot be dealt with by any existing algorithm.

In our implementation three hierarchical levels were derived based on the given search range in Figures 6 and 7. Table 1 lists the number of extracted segments at every level. Table 2 lists the average length of the extracted segments. Table 3 lists the statistics on the number of candidate matches of the extracted segments. From Table 2, the average length of segments is longer than 5 pixels at all levels for the two stereo pairs. At most levels (five of six levels as listed in Table 2), the average length of segments searched by our algorithm are greater than 10 pixels while, for other grid based image matching algorithms, the interval is only 5 pixels. This difference means that our tracking process can greatly reduce the number of matching tokens and thus saves the computation. At the coarse level, most segment paths correspond to an entire epipolar line (see Table 2). Table 3 indicates that the extracted segment paths are almost final because more than 85.3 percent of the segments have only one candidate. At the medium level, the average length of segments is greater than 10 pixels. Over 86.3 percent of the segments have no more than two candidates. This is also true at the fine level. Generally speaking, from the coarse level to the fine level, candidate segment paths dramatically increase. Because over 80 percent of segment paths at all levels tracked by our algorithm have no more than two candidates, matching ambiguities are greatly reduced. This is achieved in our algorithm by the use of global matching evidence through tracking in the similarity space, while other algorithms use every point or local peaks as their candidates in the search range. Our algorithm further reduces the computation for resolving matching ambiguities. Based on the observation that over 80 percent of segment paths at all levels have no more than two candidates, the computation of our algorithm is mainly for calculating the similarity images. The similarity image for each line in the epipolar stereo pair is not necessarily computed in whole. The computation can be divided into two parts: one used to find the tracking seeds and the other used to find the next tracking point in the small neighboring region of the current tracking point. The search range of tracking seeds is 5 pixels, while the tracking neighboring region is 3 pixels. Thus, with our algorithm, if the number of total tracking seeds is N, the sum of the length of all extracted segment paths is M pixels, and the computation of a cross-correlation coefficient is P, then the computation costs are

$$C = (5 \times N + 3 \times M) \times P. \tag{8}$$

Here P is a constant. Thus, $C(N, M)$ is a linear function. Normally, N and M increase linearly as image sizes increase. Some random variations may occur; i.e., N and M will not strictly follow the above rule. From this analysis, our algorithm can be considered as approximately computationally linear. The following further study reflects this analysis.

Fifteen different sizes of stereo pairs were input into our algorithm and the computation time was recorded. In order to

TABLE 1. NUMBER OF TOTAL EXTRACTED SEGMENTS AT EACH LEVEL

	Fine	Medium	Coarse
No. of segments in Figure 6	43,798	1,713	31
No. of segments in Figure 7	4,777	234	13

TABLE 2. THE AVERAGE LENGTH OF EXTRACTED SEGMENTS AT EACH LEVEL

		Fine	Medium	Coarse
Natural (Figure 6)	Average length (pixel)	21.2	17.3	25.5
	Percentage of image width	2.6%	10.5%	77.4%
Urban (Figure 7)	Average length (pixel)	15.6	10.4	5.5
	Percentage of image width	9.0%	29.9%	79.0%

TABLE 3. CANDIDATE MATCH STATISTICS OF EXTRACTED SEGMENTS AT EACH LEVEL

Number of candidates	Percent of extracted segments (unit: %)					
	Coarse		Medium		Fine	
	Figure 6	Figure 7	Figure 6	Figure 7	Figure 6	Figure 7
1	85.3	100.0	36.2	56.7	33.8	45.2
2	14.7	0.0	50.1	36.7	49.0	40.1
≥3	0.0	0.0	13.7	6.6	17.2	14.7

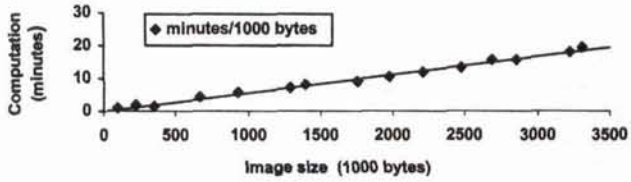

Figure 8. The relationship between the computation time and the image size.

make the running times comparable, we assigned all the test stereo pairs the same global maximum and minimum search range. A linear regression analysis between the computation time and the image size was made. Equation 9 is the result of the regression analysis: i.e.,

$$T = 0.00553 \times B. \tag{9}$$

Here T is the computation time and its unit is 1 minute, and B is the image size in units of 1000 bytes. The root-mean-square error is ± 0.63 minutes. From Figure 8, we can see that the running time fluctuates around the regression line. According to Equation 9, for a standard 23- by 23-cm frame aerial stereo pair, if the scanning resolution is 25 μm and the overlap along the flight direction is 65 percent, it will take about 28 minutes for our algorithm on a low end DEC Alpha 3000 workstation to derive the disparity map in the overlap region.

Conclusions

We propose a new image matching algorithm that tracks connected local peak points in the matching similarity space. We demonstrated the feasibility of implementing this algorithm with the surface continuity and ordering constraints during the best match tracking process. Global information available to support a local match to resolve matching ambiguities is fully utilized in such a way that unrelated global information is excluded. Thus, the solution to this problem is reliable and efficient. Our experiment indicates that our algorithm can work with a wide array of complex images.

In summary, our algorithm consists of five steps at each level:

- Extract paths by tracking connected local peak points in the similarity space;
- Divide paths sharing parts of x-coordinates into shorter paths and rearrange them into a left-to-right sequence;
- Select paths with the highest similarity values and discard those matched paths violating the surface continuity and ordering constraints;
- Fill unmatched gaps by interpolation; and
- Construct the final disparity map by filtering along the y-direction.

The algorithm can be considered as a feature-based algorithm. Such kinds of features extracted from the similarity space have not been reported in the literature. Features are normally extracted from the stereo image space. The main advantage of these types of tokens over tokens extracted from the left and right images is that the geometric distortions caused by the perspective effect between every left token and its corresponding right token is automatically eliminated. In addition, the description inconsistency caused by independent extraction of tokens in the left and right images vanishes. From the point of optimization, our algorithm may be classified into a heuristic information-based algorithm compared to information iteration propagation such as relaxation. The heuristic information is directly derived from the continuity constraint and ordering constraint and the correspondence between paths that are formed by local peaks in the similarity space and the surface profile paths in the real world. Because of the above-mentioned improvements, our algorithm is efficient and reliable and can work in a wide array of complex scenes.

Acknowledgments

This research was partially supported by the University of California Integrated Hardwood Range Management Program, a University of California DANR grant, a USFS (NCASI) grant, and a National Natural Science Foundation of China grant (49825511).

References

Ayache, N., and B. Faverjon, 1987. Efficient registration of stereo images by matching graph descriptions of edge segments, *International Journal of Computer Vision*, 1(2):107–131.

Baker, H.H., and T.O. Binford, 1981. Depth from edge and intensity based stereo, *Proceedings of 7th International Joint Conference of Artificial Intelligence*, 24–28 August, Vancouver, British Columbia, Canada, pp. 631–636.

Barnard, S.T., and M.A. Fischler, 1982. Computational stereo, *Computing Surveys*, 14(4):553–572.

Christmas, W.J., J. Kittler, and M. Petrou, 1995. Structural matching in computer vision using probabilistic relaxation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):749–764.

Cruz, J.M., G. Pajares, and J. Aranda, 1995. A neural network model in stereovision matching, *Neural-Networks*, 8(5):805–813.

Dhond, U.R., and J.K. Aggarwal, 1989. Structure from stereo—A review, *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1489–1510.

Fua, P., 1997. From multiple stereo views to multiple 3-D surfaces, *International Journal of Computer Vision*, 24(1):19–35.

Grant, I., X. Pan, and F. Romano, 1998. Neural network method applied to the stereo image correspondence problem in three-component particle image velocimetry, *Applied Optics*, 37(17):3656–3663.

Grimson, W.E.L., 1981. *From Images to Surfaces: A Computational Study of the Human Early Visual System*, MIT Press, Cambridge, Massachusetts, 274 p.

———, 1985. Computational experiments with a feature based stereo algorithm, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 7(1):17–34.

Grün, A., 1985. Adaptive least squares correlation: A powerful image matching technique, *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3):175–187.

Helava, U.V., 1988. Object-space least-squares correlation, *Photogrammetric Engineering & Remote Sensing*, 54(6):711–714.

Heipke, C., 1996. Overview of image matching techniques, *OEEPE Workshop on the Application of Digital Photogrammetric Workstations*, 04–06 March, Lausanne, Switzerland, (see *http://dgrwww.epfc.ch/PHOT/workshop/wks96/tocwks96.html*).

Hongo, S., N. Sonehara, and I. Yoroizawa, 1996. Edge-based binocular stereopsis algorithm—A matching mechanism with probabilistic feedback, *Neural-Networks*, 9(3):379–395.

Horaud, R., and T. Skordas, 1989. Stereo correspondence through feature grouping and peak cliques, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1168–1180.

Jones, G.A., 1997. Constraint, optimization, and hierarchy: Reviewing stereoscopic correspondence of complex features, *Computer Vision and Image Understanding*, 65(1):57–78.

Kang, M.S., R.H. Park, and K.H. Lee, 1994. Recovering an elevation map by stereo modeling of the aerial image sequence, *Optical Engineering*, 33:3793–3802.

Luo, A., and H. Burkhardt, 1995. An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions, *International Journal of Computer Vision*, 15(3):171–188.

Marapane, S.B., and M.T. Mohan, 1994. Multi-primitive hierarchical (MPH) stereo analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):227–240.

Medioni, G., and R. Nevatia, 1985. Segment-based stereo matching, *Computer Vision, Graphics and Image Processing*, 31(1):2–18.

Okutomi, T., and T. Kanade, 1992. A locally adaptive window for signal matching, *International Journal of Computer Vision*, 7(2):143–162.

Ohta, Y., and T. Kanade, 1985. Stereo by intra- and inter-scanline search using dynamic programming, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):139–154.

Pankanti, S., and A.K. Jain, 1995. Integrating vision modules: stereo, shading, grouping, and line labeling, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):831–842.

Rohaly, A.M., and H.R. Wilson, 1993. Nature of coarse-to-fine constraints on binocular fusion, *Journal of the Optical Society of America (Optics, Image Science, and Vision)*, 10(2):2433–2441.

Rosenholm, D., 1987. Multi-point matching using the least-squares technique for evaluation of three-dimensional models, *Photogrammetric Engineering & Remote Sensing*, 53(6):621–626.

Shapiro, L.G., and R.M. Haralick, 1987. Relational matching, *Applied Optics*, 26(10):1845–1851.

Sherman, D., and S. Peleg, 1990. Stereo by incremental matching of contours, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(11):1102–1106.

Terzopoulos, D., 1983. Multilevel computational process for visual surface reconstruction, *Computer Vision, Graphics and Image Process*, 24(1):52–96.

Venkateswar, V., and R. Chellappa, 1995. Hierarchical stereo and motion correspondence using feature groupings, *International Journal of Computer Vision*, 15(3):245–69.

Weidner, U., and W. Förstner, 1995. Towards automatic building extraction from high resolution digital elevation models, *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4):38–49.

(Received 02 August 1999; revised and accepted by January 2000; received 10 March 2000)