Multipass hierarchical stereo matching for generation of digital terrain models from aerial images

Yi-Ping Hung[1], **Chu-Song Chen**[1], **Kuan-Chung Hung**[1,2], **Yong-Sheng Chen**[1,2], **Chiou-Shann Fuh**[2]

[1] Institute of Information Science, Academia Sinica, Nankang, Taipei, Taiwan; e-mail: hung@iis.sinica.edu.tw
[2] Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan

Abstract. This paper presents a new multi-pass hierarchical stereo-matching approach for generation of digital terrain models (DTMs) from two overlapping aerial images. Our method consists of multiple passes which compute stereo matches with a coarse-to-fine and sparse-to-dense paradigm. An image pyramid is generated and used in the hierarchical stereo matching. Within each pass, the DTM is refined by using the image pyramid from the coarse to the fine level. At the coarsest level of the first pass, a global stereo-matching technique, the intra-/inter-scanline matching method, is used to generate a good initial DTM for the subsequent stereo matching. Thereafter, hierarchical block matching is applied to image locations where features are detected to refine the DTM incrementally. In the first pass, only the feature points near salient edge segments are considered in block matching. In the second pass, all the feature points are considered, and the DTM obtained from the first pass is used as the initial condition for local searching. For the passes after the second pass, 3D interactive manual editing can be incorporated into the automatic DTM refinement process whenever necessary. Experimental results have shown that our method can successfully provide accurate DTM from aerial images. The success of our approach and system has also been demonstrated with a flight simulation software.

Key words: Computer vision – Stereo matching – Surface reconstruction – DTM generation – Interactive editing

1 Introduction

Stereo vision is an important technique for recovering 3D information from 2D images. The most difficult problem in stereo vision is the stereo correspondence problem. Once the corresponding points in a stereo image pair are identified, the 3D depth can be easily computed by triangulation [10, 15, 16].

Correspondence to: Y.-P. Hung

Many stereo-matching approaches have been developed especially for digital terrain model (DTM) acquisition. Gruen [13] presented an area-based adaptive least squares correlator. The system parameters allow for local geometric image shaping and radiometric corrections, and are automatically corrected and optimized during the least square iteration. Ebner and Heipke [11] presented an approach which uses an iterative least squares adjustment to optimize the geometric and radiometric parameters simultaneously. Any desired number of images can be processed at the same time to achieve more reliable results. Schenk [27] proposed a method which determines the DTM in a coarse-to-fine approach. An image pyramid is used. In each level, the images are warped to be orthographic using the DTM obtained in the previous level. Block matching can then be applied to edge pixels on the left warped image directly, since most shape distortion is eliminated. However, all the iterative approaches described above need a good initial estimate to achieve good convergence result (typically, the initial estimate should be within three or four pixels to the true location to achieve good convergence). Kang, et al. [16] proposed an area-based stereo-matching approach for aerial sequence images. At each pixel, block matching is performed. Normalized cross-correlation is adopted as a similarity measure. The center of the search window is based on the previous matching result. However, errors may occur easily if the block matching is performed on an area with little texture. Besides, the error will propagate because the search window of the current pixel depends on previous results.

Most stereo correspondence techniques developed for DTM generation use area-based approaches. An area-based approach typically establishes stereo-matching by using a cross-correlation measure. The advantage of area-based matching is that it usually implies a strong continuity assumption [2]. However, these techniques may suffer from that (i) they are sensitive to depth variations, and (ii) they cannot easily incorporate global consistency criteria, because they attempt to resolve ambiguity based purely on local information. Hence, to overcome these problems, a preferred approach is to use a feature-based approach. For example, many approaches (not restrict to those are especially used for DTM generation) use features such as edge pixels and

zero-crossings for matching [1, 3, 12, 15, 19, 21, 23, 24, 28]. In general, feature-based approaches are more reliable than area-based approaches, because the features are more insensitive to photometric variations [18, 20].

Another popular control strategy to make stereo-matching more reliable is to use a *coarse-to-fine* scheme which can further reduce ambiguity and speed up stereo-matching [12, 22, 25, 27]. An image pyramid is usually used in a coarse-to-fine scheme. The image pyramid has been studied extensively in the past [4, 5, 26]. In coarse-to-fine stereo analysis, information extracted in a coarse level can be used to constrain the search range for matching in a finer level. For example, in [12], an edge-based stereo-matching algorithm was proposed. Edge pixels in the left and right images are matched with each other if they have the same edge polarity, approximately the same orientation, and lie on the same epipolar line. The matching process is then done using the coarse-to-fine strategy. In [19], a multiple primitive hierarchical (MPH) method was proposed. In this method, stereo analysis is performed in multiple stages, incorporating multiple primitives, utilizing a hierarchical control strategy. The primitives used include regions, linear edge segments, and edges. In a coarse-to-fine approach, the initial matching begins at a coarse scale. Then, a coarse-to-fine algorithm achieves its efficiency through iterative refinement, in which approximate results are first obtained in a lower resolution, and then are refined progressively in higher resolutions.

In this paper, we present a new multi-pass hierarchical stereo-matching approach for generation of a DTM from two overlapping aerial images. Our method is a hybrid of both the feature-based and area-based strategies, and is incorporated into a coarse-to-fine scheme. In the coarsest level of our method (i.e., in the lowest resolution image), only prominent/salient feature-primitives[1] are selected for matching. In this level, to make the matching more reliable, we use a global stereo-matching technique (the intra-/inter-scanline matching) which solves the optimal solution using dynamic programming. Hence, in the coarsest level, matching is not only based on local information, but also based on global consistency. Then, in other subsequent levels, we use area-based techniques (here, block matching is used) to refine the matching accuracy iteratively. The stereo-matching strategy developed in this paper for DTM generation is based on the following considerations:

1. Speed: notice that the size of an aerial image is usually very large (typically about 10 K × 10 K). Directly processing the highest resolution image will take too much time. Using a coarse-to-fine scheme can considerably speed up the matching process.
2. Accuracy: matching techniques based on global consistency are usually slower than those based only on local information. However, global matching techniques can usually produce more accurate matching result than local ones can. Hence, a better strategy is to use a global matching technique in the coarsest level, while using a local method in the finer levels. Compared to the approach directly using a global method for the whole

images, this strategy can save much computation time without degrading the matching performance.

The proposed approach has been used to implement an automatic DTM generation system which can make use of the results of interactive 3D editing whenever necessary. Our DTM generation system includes the following three major components.

1. Stereo images normalization. Due to the vibration and instability of the airplane during the acquisition of aerial photographs, the stereo images acquired are usually not *in correspondence*, i.e., their optical axes are not "in parallel and perpendicular to their baseline". Therefore, the epipolar lines are not exactly the horizontal scanlines of the images. To reduce the computational cost in stereo-matching, we can transform (or warp) the original stereo image pair into a *normalized* stereo image pair by using the a priori known camera parameters [14]. In the normalized stereo image pair, the corresponding horizontal scanlines are exactly the corresponding epipolar lines. Hence, the 2D stereo-matching problem can be reduced to a simpler 1D matching problem along the horizontal scanlines. For illustration, Fig. 1 shows an original stereo image pair with one image overlapping the other, and Fig. 2 shows their *normalized* stereo image pair to be used in the subsequent stereo-matching.
2. Stereo matching and DTM generation. A multi-pass hierarchical stereo-matching approach is developed to solve the stereo correspondence problem. Once the stereo correspondence of a feature point is determined, its 3D coordinates can be computed by triangulation, and finally a DTM can be obtained by interpolation. Details of our stereo-matching approach are described in this paper.
3. Interactive 3D editing. Since it is generally impossible to obtain faultless stereo correspondence with state-of-the-art stereo-matching technologies, implementing a 3D interactive environment for inspecting and editing the DTM becomes crucial and useful in acquiring accurate DTMs. In our system, we use a pair of stereo glasses from StereoGraphics Co. (or more precisely, CrystalEyes stereo glasses connected with an SGI Indigo[2]) to allow the user to perceive the 3D terrain directly from the aerial images, and then to compare it with the overlaid DTM generated by the computer. If there is any inconsistency between them, the user can operate a 3D mouse (here we use a Microscribe-3DX having six degrees of freedom) to edit the DTM accordingly. More details of the interactive 3D editing process will be presented in Sect. 5. The modified DTM can then be sent to the hierarchical stereo-matching process for further refinement of the DTM using the pyramids of aerial images.

This paper is organized as follows. Section 2 gives an overview of the multi-pass hierarchical stereo-matching system. Section 3 describes the intra-/inter-scanline matching method used in the coarsest level of the first pass. Section 4 introduces the hierarchical block-matching method used in our system. Section 5 describes our interactive 3D editing environment. Section 6 shows some experimental results, and Sect. 7 gives some conclusions.

---

[1] Here, we use sufficient long edge curves which do not lie along the epipolar lines, as described in Sect. 3.
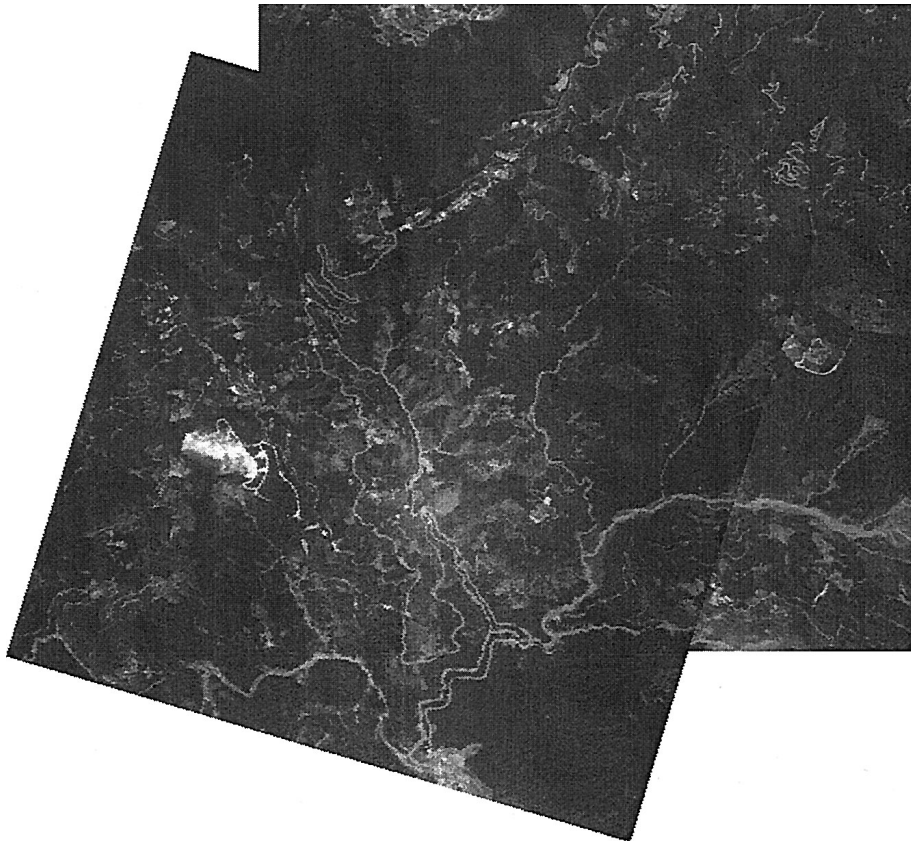
Fig. 1. An original stereo image pair
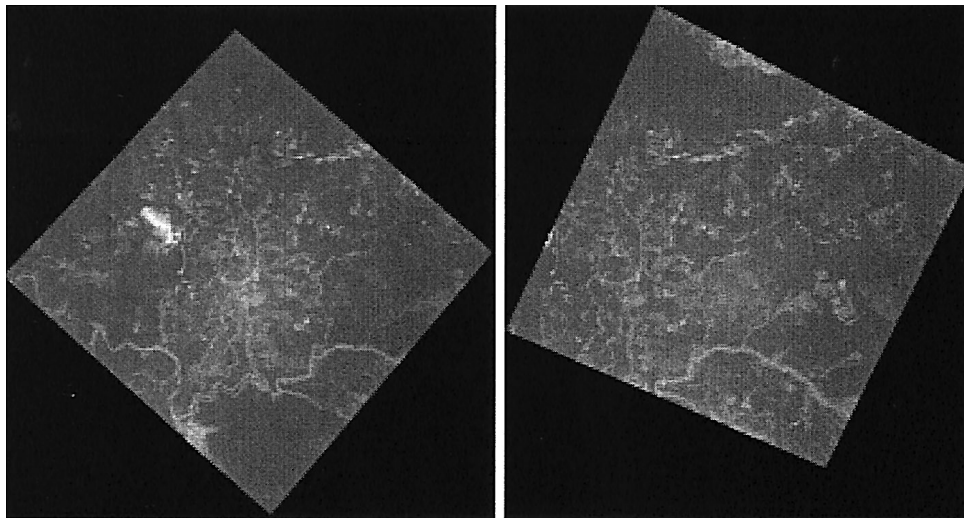
Fig. 2. The normalized image pair of the images shown in Fig. 1, where the corresponding horizontal scanlines are exactly the corresponding epipolar lines
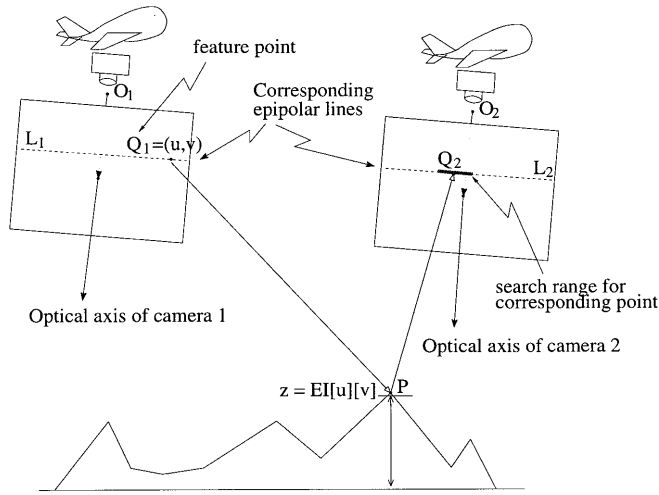
2 Overview of the multi-Pass hierarchical stereo matching

Figure 3 illustrates the data flow of the first three passes in our multi-pass hierarchical stereo-matching method. When considering the process from Pass 1 to Pass 2, our method is a sparse-to-dense method in terms of the 2D features used and the 3D features obtained. This sparse-to-dense paradigm increases the matching correctness in a systematic way. In Pass 1, the matching will be performed only at the image locations near the salient features; in our system, the matching is only performed in the neighboring regions of long edge segments. The 3D feature points generated by Pass 1 are relatively sparse; however, they provide a good initial condition for Pass 2. In Pass 2, instead of using only the features near long edge segments, all high-contrast blocks will be used for stereo matching. Hence, the 3D features generated by Pass 2 will be denser and distributed over the whole image.
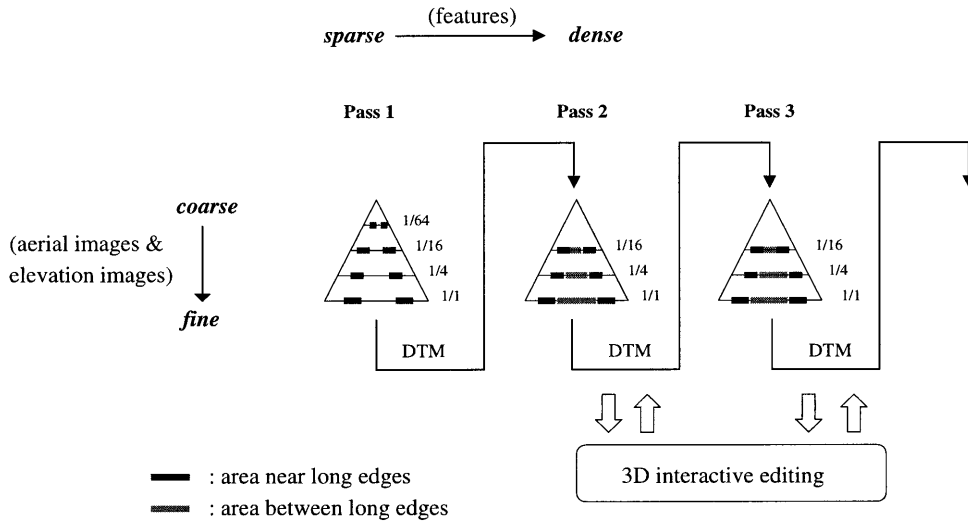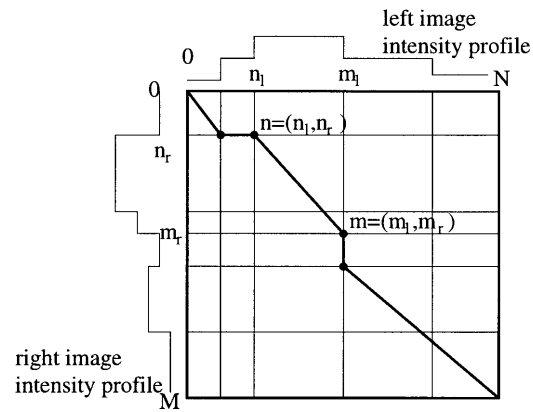
Fig. 3. Illustration of the first three passes of the multi-pass hierarchical stereo-matching approach
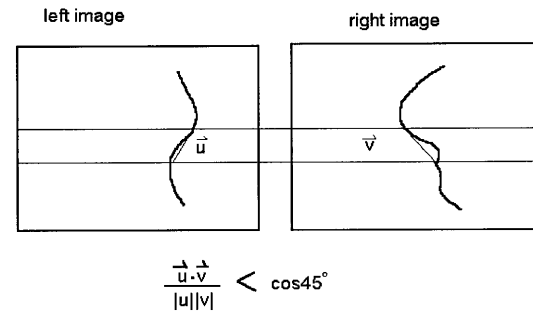

Fig. 4. The elevation image and its use in finding correspondences

Fig. 5. The intra-scanline search

$$\frac{\vec{u} \cdot \vec{v}}{|u||v|} < \cos 45°$$

Fig. 6. A constraint for removing unreliable matches

Furthermore, each pass (sparse or dense) is performed with a coarse-to-fine matching strategy, as illustrated in Fig. 3. An image pyramid is generated to perform the coarse-to-fine matching. The size of an aerial image is usually very large (typically about 10 K × 10 K). Hence, it is time-consuming and problematic to find correspondences directly with such a high-resolution image. In our approach, stereo correspondences in a coarser level are first determined by using a low-resolution image; then, correspondences in the subsequent finer levels can be refined by using higher resolution images.

In principle, stereo-matching methods can be divided into two classes, *local* methods and *global* methods. In a local method, the correspondences are determined by exploiting only local similarities. On the other hand, in a global method, correspondences are determined in a global way. Generally speaking, local methods are faster than global ones, but usually produce more incorrect matches than the global ones do. In our system, we first use a global method (i.e., the intra-/inter-scanline matching method [8, 23]) to find a good initial DTM in the coarsest level of Pass 1, and then use a local method (i.e., the hierarchical block-matching method) to refine the DTM in the subsequent finer levels of Pass 1. For Pass 2 and the following passes, only the block-matching method are adopted in order to save computation time (of course, without degrading the matching performance).

There are two main terrain data structures used in our system. One is the *elevation image* (EI), and the other is the *digital elevation map* (DEM). An EI contains a 2D array

of terrain elevation data described with respect to the left image. Notice that, while the EI is uniformly indexed on the x-y plane of the left camera coordinate system, they are not uniformly indexed in the world coordinate system. Every pixel of EI stores the 3D elevation value of the pixel on the left image. The resolution of EI varies for different layers in an image pyramid. Given a set of feature points in the left image, it is easy to obtain the corresponding image coordinates on the right image using a given EI and the camera parameters associated with the stereo image pairs (refer to Fig. 4 and [7]).

A DEM is a terrain elevation map described with respect to the world coordinate system. In our system, the DEM is uniformly indexed in the world coordinate system (but not in the image coordinate system) and is used as the final output for displaying the terrain model.

Both the EI and DEM store 3D data in a 2D array. While the DEM is better used for output, the EI is more suitable for stereo-matching. Since our stereo-matching is based on relatively sparse feature points, there are tremendous amounts of pixels in the EI (or DEM) that contain no direct 3D data. For those pixels, 3D data have to be computed using interpolation. Interpolation is also used between the conversion of EI and DEM. Since the EI is not uniformly indexed with respect to the world coordinate system and it usually has a different resolution from the DEM, there can be missing grid points when converting from an EI to a DEM, or vice versa. Therefore, interpolation has also to be done to fill those missing grid points.

3 Intra-/inter-scanline matching

The intra-/inter-scanline matching method can be divided into two phases [23]. The first phase is the *intra-scanline search* phase, and the second phase is the *inter-scanline consistency* phase.

The intra-scanline search determines the corresponding edge points [6] between a scanline pair. It can be viewed as a path-finding problem on a 2D search plane. As shown in Fig. 5, the horizontal axis is associated with the intensity profile on the left scanline, and the vertical axis is associated with the intensity profile on the right scanline. The vertical lines represent the positions of edge points on the left scanline, and the horizontal lines represent the positions of edge points on the right scanline. The intersections of the vertical and horizontal lines can be regarded as the possible matching edge pairs between the two scanlines. Each intersection is referred to as a node. Now, the problem of finding correspondences is transformed into a problem of finding an optimal path on the 2D search plane. The definition of path cost and the algorithm for finding optimal path is the same as that introduced in [8].

In the intra-scanline search, each scanline pair is processed independently. However, for an edge extending across scanlines, its stereo correspondence in one scanline should have a strong dependency on the correspondences in the neighboring scanlines. Edge-based inter-scanline match-consistency is used to refine the results obtained by the intra-scanline search. Here, the principle used for inter-scanline
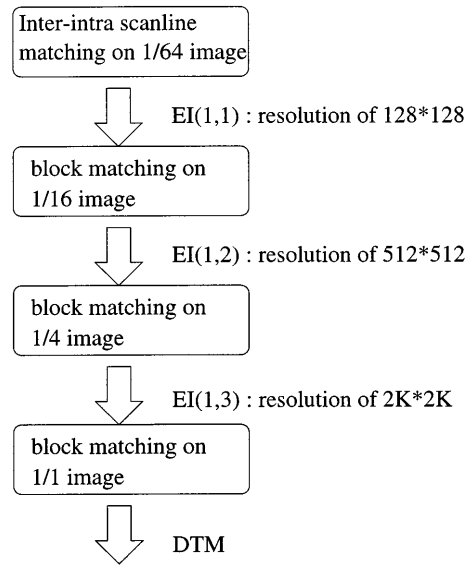

Fig. 7. The flow chart of Pass 1. Here, $EI(i, j)$ denotes the EI generated at level j of Pass i

match-consistency is a "winner-take-all" scheme, same as the one introduced in [8].

To eliminate unreliable match, short edges are disregarded in the first place and only long-enough edges are used for matching. To improve the correctness of matching, we propose a useful constraint to remove unreliable matches: if the direction difference between two matching edges is too large, this match is viewed as an unreliable match and will be removed (see Fig. 6). In our experience, this constraint can considerably remove incorrect matches, while preserving the correct ones for the intra-scanline search. In addition, unlike the indoor stereo image pairs, the aerial stereo image pairs seldom contain stereo occlusion parts, at least in the coarsest level. Hence, we do not have to consider stereo occlusion in stereo-matching of aerial images. In fact, this strategy can also make the intra-/inter-scanline matching more efficient and reliable since the definition of the cost for stereo occlusion is usually very ad hoc for most stereo-matching methods.

4 Hierarchical block matching

The rough DTM obtained with the intra-/inter-scanline matching method described in the last section can be used as an initial condition for the hierarchical block matching described in this section. In our current implementation, block matching is used in the 1/16, 1/4, and 1/1 resolution levels of Pass 1 and all the following processing after Pass 1 (refer to Fig. 3).

In this work, block matching is performed when we have a rough DTM obtained either from the upper layer or from the previous pass. Therefore, the search region will be restricted within a local range. In addition, the search can be performed in a nearly 1D region along the epipolar line.

The following cross-correlation measure is used as the similarity measure for block matching in our system:

$$\rho \equiv \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(A_{ij} - \bar{M}(A))(B_{ij} - \bar{M}(B))}{\sqrt{var(A)} \sqrt{var(B)}}, \qquad (1)$$

where A_{ij} is the intensity of the (i,j)th pixel in block A, $\bar{M}(A)$ and $var(A)$ are respectively the mean value and the variance of block A, and the number of pixels in a block is mn.

The advantage of using cross-correlation is that the cross-correlation is invariant to linear brightness changes between two blocks. Since, in general, the intensity values of the same terrain point in two aerial images may be different (because the two aerial images are taken at different times and positions), using cross-correlation can provide more stable results if the brightness change between the two corresponding blocks can be locally approximated by a linear function.

4.1 Block matching for Pass 1

The flow chart of Pass 1 is shown in Fig. 7. For a specific resolution level, the general procedure can be divided into the following three steps:

1. Refine the matching point pair. Assume that (p_l, p_r) is a pair of matching points obtained from the previous level (where p_l is in the left image, and p_r is in the right image). If this match is correct, the error of the edge position should be within one pixel in the previous level. Because the resolution is four times larger for this finer level, we set an area centered at p_l with size of ± 4 pixels. In this area, a refined edge point is set to be the one having the largest contrast (e.g., using the Sobel operator). Then, to find a more accurate corresponding point, block matching is performed on the right image in the ± 4 pixel region centered at p_r. After this step, both the feature and the matching positions obtained from the previous level have been refined to be more accurate.

2. Extract features near the refined edge point. To extract more feature points to be used in the block matching in the next step, we use the EJO (early jump-out) method [9] within a pre-specified area centered at the refined edge point.

3. Acquire more matching point pairs. For every feature point extracted in the previous step, we compute an initial correspondence position on the right image based on the current estimate of DTM (refer to Fig. 4 and [7]). Then, within a search region centered at this initial correspondence position, the optimal matching point is chosen to be the image location that gives the largest cross-correlation measure in the block-matching process.

4. Verification. To increase the reliability of the stereo-matching result, we use the following process for further verification.

If the cross-correlation measure of the optimal match is smaller than a threshold, then this match is viewed as an unreliable one and will be discarded (i.e. , it will not be used as a primitive for interpolation).

Hence, only the highly correlated match will be accepted in this step. According to our experience, if a matching found is not correct at the coarser resolution, then it is very likely to be removed by the verification step in a subsequent finer level.
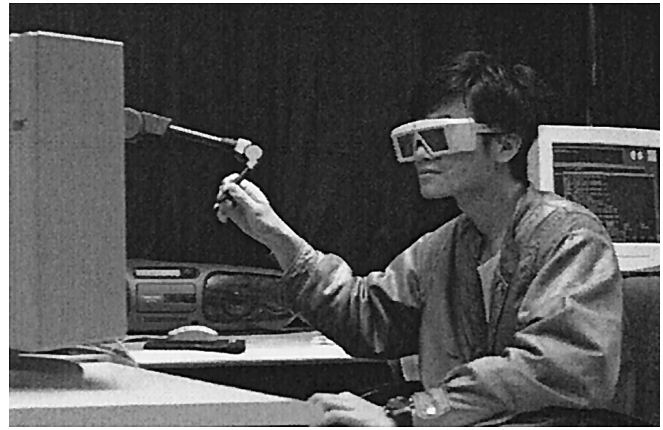
Fig. 8. A picture illustrating that an operator is using the interactive 3D editing system which includes a pair of stereo glasses (here we use a CrystalEyes stereo glasses connected with a SGI Indigo^2) and a 3D mouse (here we use a Microscribe-3DX having six degrees of freedom)

The DTM generated by Pass 1 is interpolated from a set of relatively sparse 3D points near long-enough edge segments. This DTM will be sent to Pass 2 to compute a more accurate terrain model.

4.2 Block matching for Pass 2

The EI generated by Pass 1 is used to generate initial correspondences for Pass 2. The procedure of block matching in Pass 2 is similar to that in Pass 1. The main difference is that, at each level of Pass 2, step 2 will extract feature points from the whole image, instead of from the regions neighboring long-enough edge segments.

4.3 Block matching for Pass 3

The DTM generated from Pass 2 of our system is already quite accurate for most of the case (the evaluation of the matching accuracy is based on a qualitative verification by operators as described in the next section). In Pass 3, the DTM can be further manually edited (also described in the next section) and then be used as a better initial starting point for stereo-matching. This human-aided semiautomatic refinement can be repeated until the operator is satisfied with the resulting DTM.

5 Interactive 3D editing system

As mentioned in Sect. 1, our interactive 3D editing system includes two major devices: a pair of stereo glasses and a 3D mouse, as shown in Fig. 8. A stereo aerial image pair is displayed on the screen with some disparity, and the two images are switched at a high frequency (about 120 Hz) and synchronized with the stereo glasses. In this way, the human
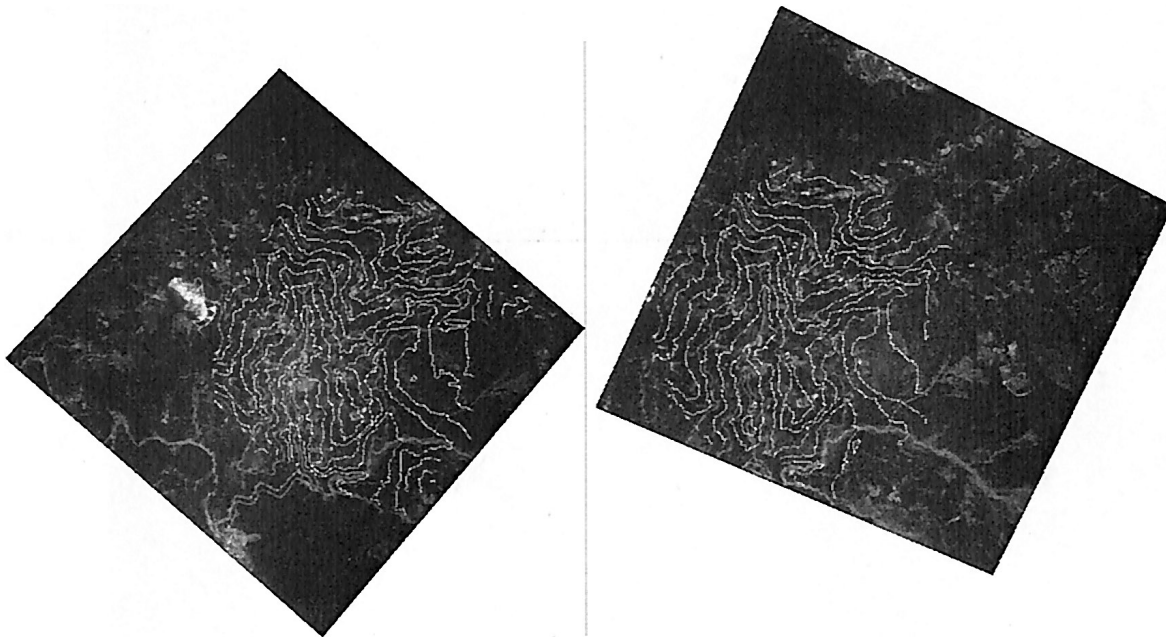
Fig. 9. The stereo pair of the aerial images shown in Fig. 2 overlaid with the computed constant elevation contours. The operator wearing a pair of stereo glasses can then perceive the difference between the computed DTM and the real 3D terrain formed naturally by the inherent capability of human stereo perception
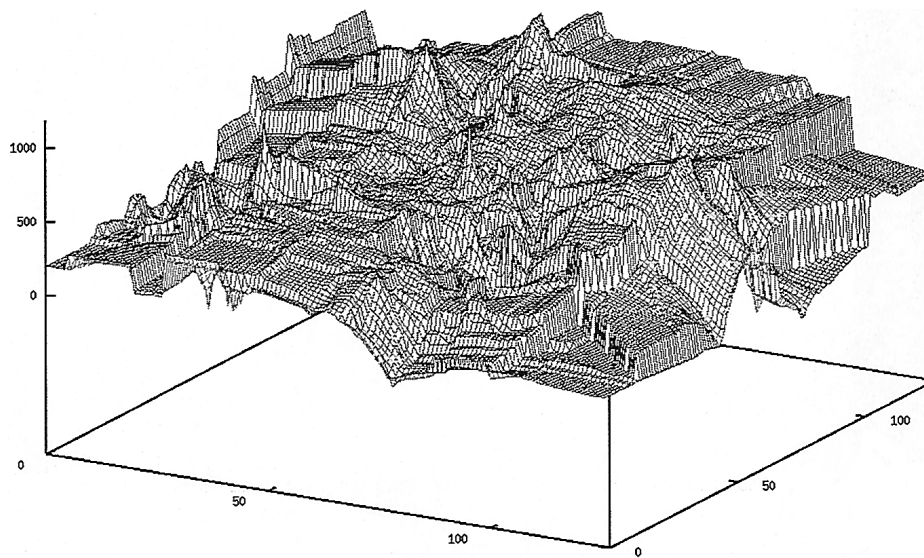
Fig. 10. The DTM generated by using the intra-/inter-scanline matching method in Pass 1 with the images of 1/64 resolution

brain can register a sequence of left- and right-eye images and fuses the two images by stereopsis. Hence, the human operator will be able to perceive the 3D terrain in a natural and comfortable way.

In stereo vision, an important problem is to evaluate the accuracy of stereo-matching. An objective evaluation requires the comparison of the generated 3D data with the *ground truth*. However, the problem of obtaining ground truth is difficult, and a simpler way is to provide some qualitative justification for the generated DTM. To justify our results qualitatively, we have developed a stereo display system which allows the human operator to perceive and evaluate the accuracy of stereo-matching directly in the 3D

space. By incorporating the human-stereo-perception-based evaluation system with an interactive 3D editing mode, a human operator is allowed to edit the incorrect parts of the generated DTM directly in the 3D space, as described in the following.

To allow the human operator to compare the inconsistency between the perceived terrain depth and the DTM generated after Pass 2, we first generate some constant elevation contours from the computed DTM. By projecting and overlaying those 3D contours onto the left and right aerial images as shown in Fig. 9, the operator wearing a pair of stereo glasses can then perceive the difference between the computed DTM and the real 3D terrain formed naturally by the
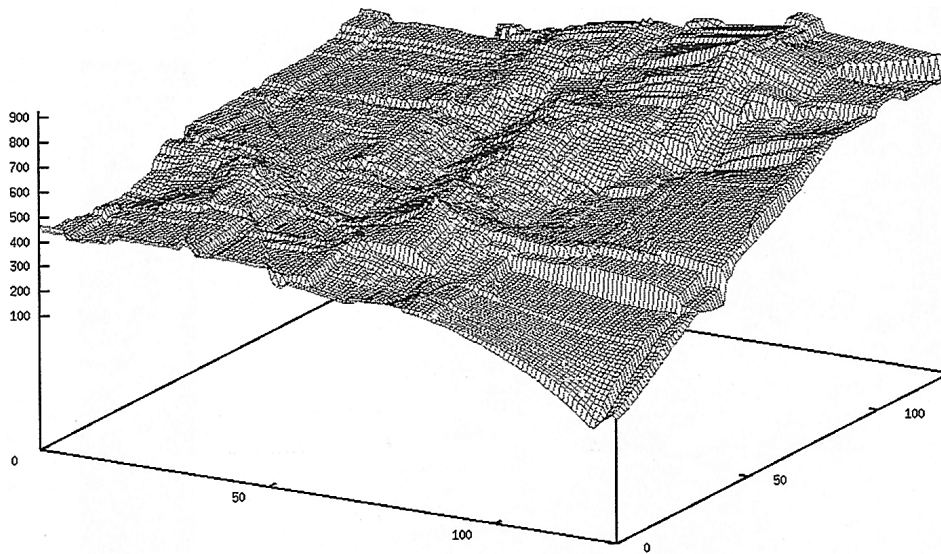
Fig. 11. The DTM generated by using the block matching in Pass 1 with the images of 1/16 resolution
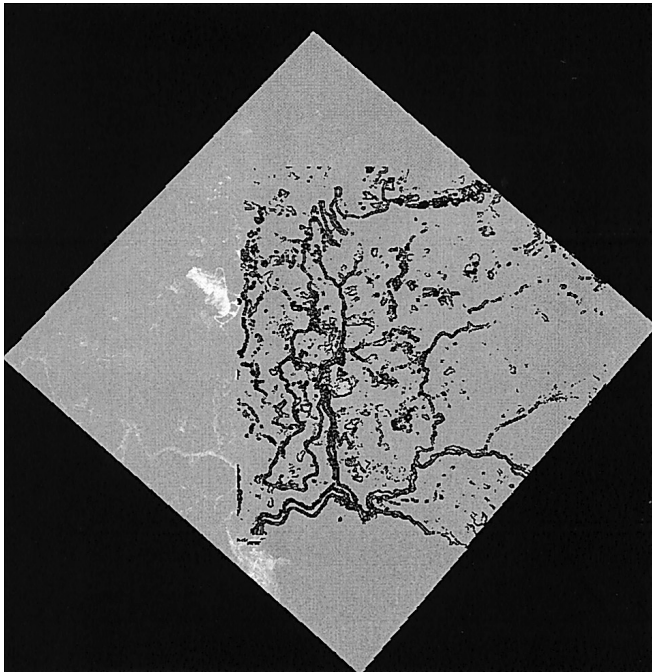
Fig. 12. The feature points extracted in the left image of 1/16 resolution in Pass 2

inherent capability of human stereo perception. If the constant elevation contours perceived in 3D space attach well to the perceived 3D terrain, then it implies that the computed stereo-matching is accurate. Otherwise, the operator can use the 3D mouse to "drag" the incorrect part to a roughly correct position, and generate a human-corrected DTM. The human-corrected DTM then serves as a better initial starting point and is sent to Pass 3 to obtain a more accurate DTM.

In fact, the greater part of the DTM generated by Pass 2 has already aligned quite well when observing through the stereo-glasses. According to our experience, only a little human assistance is needed to fix the incorrectly matched portion (whose ratio is usually less than 5%).

6 Experimental results

The size of the aerial images we use here is about 8 K × 8 K. After applying the intra-/inter-scanline matching method to the 1/64 level (i.e., the coarsest level) of Pass 1, a DTM can be obtained by using interpolation and extrapolation, which is shown in Fig. 10. It can be seen that this DTM is very rough and has quite a few bumping noises. This DTM was sent to the 1/16 level of Pass 1 as an initial starting point for stereo-matching using local search. Figure 11 shows the DTM obtained after the processing in the 1/16 level of Pass 1. Most of the bumping noises in Fig. 10 have been removed, since only those matches that are highly correlated and pass the verification step in the 1/16 level are preserved. That is, the outliers generated by the intra-/inter-scanline search have been removed by the verification step of the finer level. This new DTM is then sent to the even finer levels (i.e., the 1/4 level and the 1/1 level) for further refinement.

In Pass 1, only those feature points near long edge segments are used for block matching. In Pass 2, all feature points are used for block matching. Figure 12 shows the feature points extracted in the left image of 1/16 resolution in Pass 2. Figure 13 shows the final output DTM of Pass 2. If we plot the contours of constant elevation on top of the aerial images and wear the stereo glasses to view the 3D stereo images, we can see that this generated DTM is correct almost everywhere, except for a small area in the middle-right part of the left image shown in Fig. 2. It is mainly because there are no salient long edge segments in this part of image, thus the intra-/inter-scanline matching of the coarsest level did not provide a good enough initial estimate for the subsequent local searching steps in the finer level. Based on some human-aided modifications with our interactive 3D editing system, stereo-matching in Pass 3 can produce an even better DTM, as shown in Fig. 14. By comparing Fig. 14 with Fig. 13, it can be observed that a small hill has been pulled up in the middle-right part of the DTM.

Once the DTM is constructed, the computer can render different perspective views of the 3D terrain by mapping the texture of the aerial images onto the DTM. For exam-
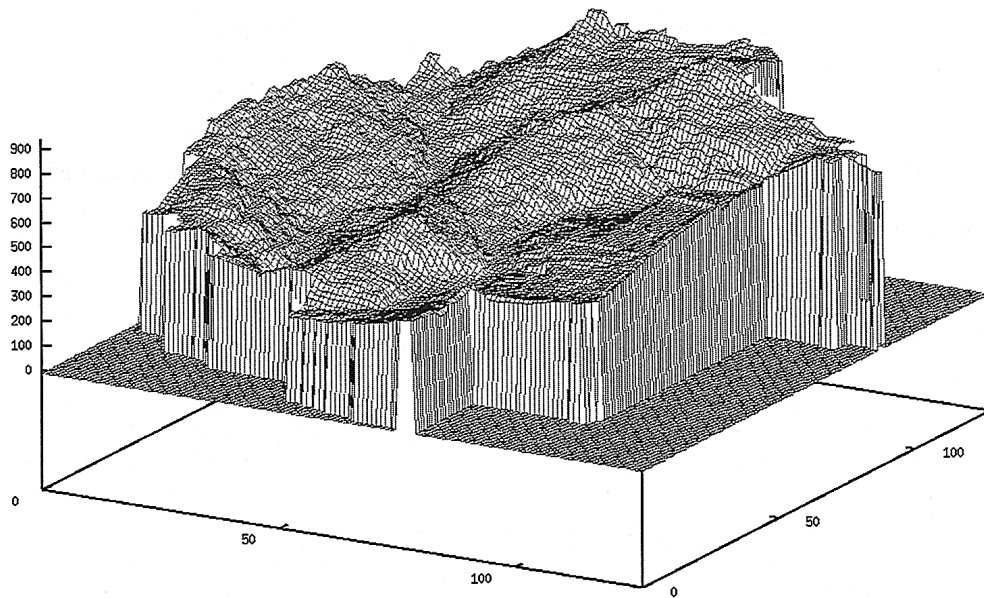
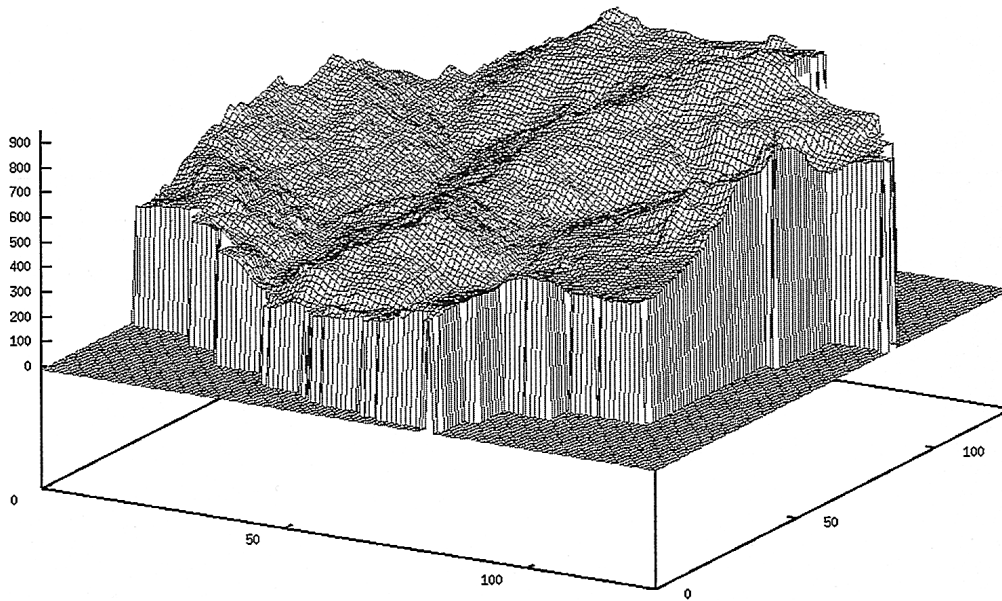**Fig. 13.** The output DTM of Pass 2

Fig. 14. The output DTM of Pass 3

ple, Fig. 15 is a perspective view generated by a computer-simulated camera located at an arbitrarily given viewpoint, and Fig. 16 shows the part of the aerial image used for rendering Fig. 15. The image patch in Fig. 16 is the central portion of the left aerial image shown in Fig. 1, and covers only about 1/16 of the original image.

Figure 17 shows another stereo pair of aerial images covering a terrain crossed by a river. The DTM obtained with our system is shown in Fig. 18, which is also very accurate (according to the qualitative verification described in Sect. 5).

7 Conclusions

In this paper, we have proposed a multi-pass hierarchical method for stereo-matching. This method has been tested with the DTM generation using aerial images. As with most iterative approaches, the choice of a good initial condition is an important issue here. Our solution is to use an effective global matching technique – the intra-/inter-scanline matching method for determining a good initial match at the coarsest level of the first pass. Based on the result obtained with this global matching at the coarsest level, a multi-pass hierarchical block-matching approach is used to refine the DTM in a systematic way.
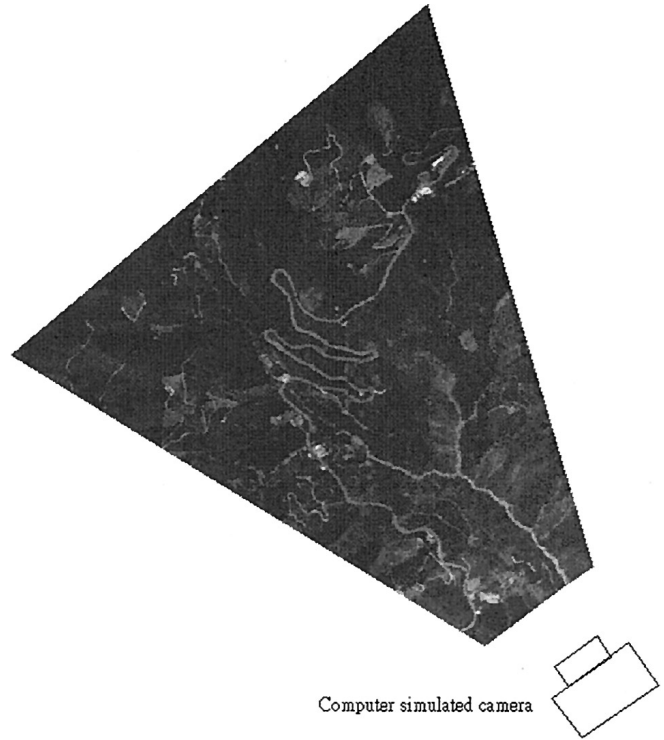
Typically, a multi-scale stereo-matching method is performed from coarse to fine in just one single pass. A significant characteristic of our method is that multiple passes

Fig. 15. A perspective view generated by a computer simulated camera located at an arbitrarily given viewpoint

Computer simulated camera

Fig. 16. A central portion of the left aerial image in Fig. 1, which is used to generate the computer-generated view shown in Fig. 15

are used to improve the matching performance. In the first pass, our system uses only the feature points that are close to long-enough edge segments. The multi-pass architecture is critical, because the initial condition interpolated from the 3D data obtained by the global matching method is a better initial condition for those features closer to the long edge segments than for those farther away features. Some characteristics of this work are described below.

1. We have developed a multi-pass coarse-to-fine approach for DTM generation, which is a hybrid of both the feature-based and area-based strategies. In the coarsest level of the first pass, we use a global matching technique to match prominent/salient feature primitives. In the subsequent finer levels of the first pass, we use local area-based methods to iteratively refine the generated DTM. To generate more accurate DTM, we use more features (i.e., increase the feature density) in the second pass. Human-aided corrections are performed in the third pass and the following passes if needed. Experimental results have shown that this multi-pass hierarchical approach can automatically produce good DTMs.
2. By overlaying constant elevation contours to the stereo pair of aerial images, our system allows the user to evaluate the stereo-matching accuracy by exploiting the inherent capability of human stereo perception.

3. By incorporating the human-stereo-perception-based evaluation system with an interactive 3D editing mode, our system allows the human operator to edit the incorrect parts of the generated DTM directly in the 3D space. To get a more accurate DTM, the results obtained by the interactive 3D editing can then be served as a better initial estimate for the stereo-matching algorithm.

In summary, we have successfully designed and implemented a system for automatic generation of DTM from aerial images. The core of this system is the multi-pass hierarchical stereo-matching method described in the paper. Based on the DTM obtained with our system, we have also implemented a flight simulation software with texture-mapping techniques. We are currently working on extending this stereovision method to 3D model reconstruction using close-range images and indoor images.
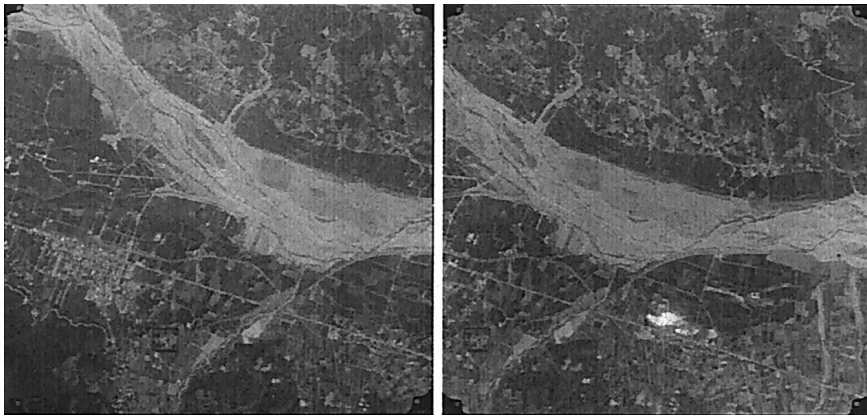
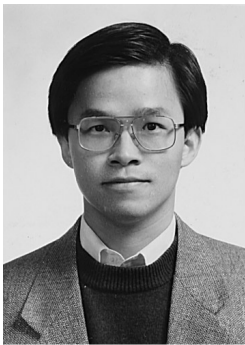**Fig. 17.** A stereo image pair of a terrain crossed by a river
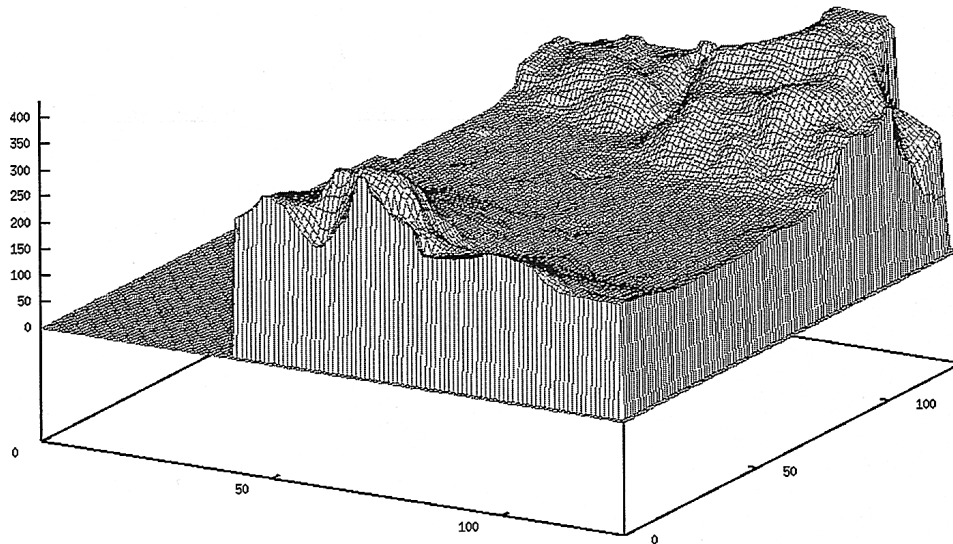
Fig. 18. The output DTM using the aerial image pair shown in Fig. 17

References

1. Baker HH, Binford TO (1981) Depth from Edge and Intensity Based Stereo. In: Proc. 7th Int. Joint Conf. Artificial Intelligence, pp 631–636
2. Barnard ST, Fischler MA (1982) Computational Stereo, ACM Comput Surv 14: 553–572
3. Boyer KL, Kak AC (1988) Structural Stereopsis for 3-D Vision. IEEE Trans Pattern Anal Mach Intell 10: 144–166
4. Burt PJ (1981) Fast Filter Transforms for Image Processing. Comput Graphics Image Process 16: 20–51
5. Burt PJ (1988) Smart Sensing within an Pyramid Vision Machine. Proc. IEEE 76: 1006–1013
6. Canny J (1986) A Computational Approach to Edge Detection. IEEE Trans Pattern Anal Mach Intell 8: 679–698
7. Cernuschi-Frias B, Cooper DB, Hung YP, Belhumeur PN (1989) Toward a Model-Based Bayesian Theory for Estimating and Recognizing Parameterized 3-D Objects Using Two or More Images Taken from Different Positions. IEEE Trans Pattern Anal Mach Intell 11: 1028–1052
8. Chen CS, Hung YP, Chiang CC, Wu JL (1997) Range Data Acquisition Using Color Structured Lighting and Stereo Vision. Image Vision Comput 15: 445–456
9. Cooper J, Venkatesh S, Kitchen L (1993) Early Jump-Out Corner Detectors. IEEE Trans Pattern Anal Mach Intell 15: 823–828
10. Day T, Muller JP (1989) Digital Elevation Model Production by Stereo-Matching SPOT Image-Pairs: A Comparison of Algorithms. Image Vision Comput 7: 95–101
11. Ebner H, Heipke C (1988) Integration of Digital Image Matching and Object Surface Reconstruction. Int Arch Photogrammetry Remote Sensing 27: III-534–III-545
12. Grimson WEL (1985) Computational Experiments with a Feature Based Stereo Algorithm. IEEE Trans Pattern Anal Mach Intell 7: 17–34
13. Gruen A, Baltsavias E (1986) High Precision Image Matching for Digital Terrain Model. Int Arch Photogrammetry Remote Sensing 26: III-284–III-296
14. Haralick RM, Shapiro LG (1993) Computer and Robot Vision, Vol. 2, Addison-Wesley, Reading, Mass.
15. Hsieh YC, Mckeown DM Jr, Perlant FP (1992) Performance Evaluation of Scene Registration and Stereo Matching for Cartographic Feature Extraction. IEEE Trans Pattern Anal Mach Intell 14: 214–238
16. Kang MS, Park RH, Lee KH (1994) Recovering an Elevation Map by Stereo Modeling of the Aerial Image Sequence. Opt Eng 33: 3793–3802
17. Lim HS, Binford TO (1987) Structural Correspondence in Stereo Vision. In: Proc. Image Understanding Workshop, Feburary 1987. Defence Advanced Research Projects Agency, USA, pp 234–241

18. Marapane SB, Trivedi MM (1989) Region-Based Stereo Analysis for Robotic Applications. IEEE Trans Syst Man Cybern (Special Issue on Computer Vision) 19: 1447-1464
19. Marapane SB, Trivedi MM (1994) Multi-Primitive Hierarchical (MPH) Stereo Analysis. IEEE Trans Pattern Anal Mach Intell 16: 227–240
20. Medioni G, Nevatia R (1985) Segment-Based Stereo Matching. Comput Vision Graphics Image Process 31: 2–18
21. Nasrabadi NM, Liu Y (1989) Stereo Vision Correspondence Using a Multi-channel Graph Matching Technique. Image Vision Comput 7: 237–245
22. Nishihara HK (1984) Practical Real-Time Imaging Stereo Matcher. Opt Eng 23: 536–545
23. Ohta Y, Kanade T (1985) Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming. IEEE Trans Pattern Anal Mach Intell 7: 139–154
24. Pong T, Haralick RM, Shapiro LG (1989) Matching Topographic Structures in Stereo Vision. Pattern Recogn Lett 9: 127–136
25. Quam LH (1984) Hierarchical Warp Stereo. In: Proc. DARPA Image Understanding Workshop, New Orleans, La. Defence Advanced Research Projects Agency, USA, pp 149–155
26. Rosenfeld A, (ed) (1984) Multiresolution Image Processing and Analysis. Springer, New York
27. Schenk T (1989) Determination of DEM Using Iteratively Rectified Images. Tech Notes Photogramm 7: 1–12
28. Watanabe M, Ohta Y (1990) Cooperative Integration of Multiple Stereo Algorithms. In: Proc. 3rd Int. Conf. Computer Vision, Osaka, Japan. IEEE Computer Society Press, pp 476–480

Kuan-Chung Hung received a B.S in 1995 and an M.S. in 1997, respectively, both from the Department of Computer Science and Information Engineering, National Taiwan University. He is now in the army. His research interests inlcude 3D modeling and virtual reality.

Yong-Sheng Chen received the B.S. degree in computer science from the National Chiao-Tung University, Hsing-Chu, Taiwan, in 1993, and the M.S. degree in computer engineering from the National Taiwan University, Taipei, Taiwan, in 1995. He is now a research assistant at the Institute of Information Science, Taiwan, and a Ph.D. student at the National Taiwan University. His research interests include computer vision, virtual reality, and human-machine interaction.
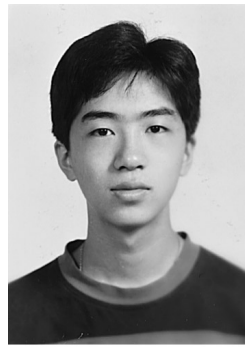

Yi-Ping Hung received a B.S. in electrical engineering from the National Taiwan University in 1982. He received an M.S. from the Division of Engineering, an M.S. from the Division of Applied Mathematics, and a Ph.D. from the Division of Engineering at Brown University, USA in 1987, 1988 and 1990, respectively. He joined the Institute of Information Science at Academia Sinica as an associate research fellow in February 1990 and became a research fellow in 1997. He is also an adjunct professor in the Department of Computer Science and Information Engineering at the National Taiwan University. His research interest includes computer vision, pattern recognition, image processing, computer graphics, virtual reality, artificial intelligence and robotics.

Chiou-Shann Fuh received the BS degree in computer science and information engineering from National Taiwan University, Taipei, Taiwan, in 1983, the M.S. degree in computer science from Pennsylvania State University, University Park, Pa, in 1987, and the Ph.D. degree in computer science from Harvard University, Cambridge, Mass., in 1992. He was with AT&T Bell Laboratories and engaged in performance monitoring of switching networks from 1992 to 1993. Since 1993, he has been an associate professor in the Computer Science and Information Engineering Department at the National Taiwan University, Taipei, Taiwan. His current research interests include digital image processing, computer vision, pattern recognition, and mathematical morphology.

Chu-Song Chen received a B.S. in Control Engineering from the National Chiao-Tung University, Hsing-Chu, Taiwan, in 1989. He received an M.S. in 1991 and a Ph.D in 1996, respectively, both from the Department of Computer Science and Information Engineering, National Taiwan University. He is now a postdoctoral fellow of the Institute of Information Science, Academia Sinica. His research interests include 3D computer vision, virtual reality, and machine learnig.