Automatic DEM generation by multi-image feature based matching

Hans-Gerd Maas
Institute of Geodesy and Photogrammetry
ETH - Hoenggerberg, CH - 8093 Zurich, Switzerland
Phone +41 - 1 - 633 3058, Fax +41 - 1 - 633 1101, e-mail gerd@geod.ethz.ch

Commission III

Keywords: digital terrain model, feature based matching, epipolar lines, multi-image

ABSTRACT:

One of the major difficulties in the automatic generation of digital elevation models (DEMs) is the procurement of approximate values. This presentation shows a new development based on the extraction of discrete points by an interest-operator and epipolar line intersection techniques in multiple overlapping images. The method can be considered an extension of well-known automatic DEM generation approaches from a stereo technique to a multi-image technique, and it solves the problem of provision of approximate values inherently. In contrast to most stereo-based techniques, the approach is not based on image pyramids, but on the consequent exploitation of the geometric strength of multiple images, implemented via the intersection of epipolar lines. Although not outlined for a complete DEM generation yet, the method may be very valuable for a hypothesis-free generation of good approximate values for other (e.g. area-based) DEM generation techniques or for the refinement of automatically generated DEMs degraded by smoothing effects.

Applications of the technique to DEM generation in mountain regions based on scanned overlapping imagery of a regular block with 80%/60% overlap and on stillvideo imagery taken from a helicopter showed good results and blunder rates of 0 - 0.5% without introducing assumptions on the maximum steepness of the terrain and without any post processing.

INTRODUCTION

The automatic generation of digital elevation models (DEM) has gained large attention among photogrammetrists in the past 10-15 years. A wide variety of approaches has been developed and presented in the literature, and automatic DEM generation packages are meanwhile commercially available in several digital photogrammetric workstations. A very rough classification of methods distinguishes between area based techniques computing correlations between greyvalue patches of images and feature based techniques establishing correspondences between extracted features of images. Techniques based on least-squares-matching (Gruen, 1985) are sometimes considered a combination of both.

A critical point in all automatic DEM generation methods is - besides the recognition (and avoidance) of objects above the terrain surface like houses or trees - the procurement of approximate values in regions where no pre-knowledge on the terrain surface is given. Most methods employ image pyramid techniques or interactively selected starting points to solve this problem:

- Image pyramid: in a coarse-to-fine approach DEMs are generated at successively finer resolution levels of the images, starting with a horizontal plane as approximation for the highest level of the pyramid; this approach is e.g. implemented in the commercial DEM generation package MATCH-T (Krzystek, 1991) and was also used by (Baltsavias/Stallmann, 1992).
- Interactive starting points: an operator sets one (or a few) starting points interactively, and the matching procedure moves into all directions from the starting point(s), assuming a certain maximum terrain slope. This approach is e.g. being used in the industrial surface measurement package INDUSURF (Schewe, 1987).

Both techniques work well in most cases, but may show problems in regions with steep surface gradients, occlusions or certain frequency patterns in the image texture. Especially in area based methods the steps in image pyramids are sometimes too big, so that approximate values cannot be considered sufficient for the consecutive pyramid level and points get lost.

The method presented in this paper does not require either one of these techniques for generating approximate values, but provides a solution which is independent on the availability of approximate values or pre-knowledge on the maximum terrain slopes; only very rough pre-knowledge (like estimated minimum and maximum terrain height for the whole observation region) is required. The procedure is similar to the procedure followed by MATCH-T: Discrete points in the images are extracted by an interest operator (Foerstner, 1986), and correspondences between points are established using epipolar lines. These epipolar lines may become rather long when the terrain shows large height differences, leading to multiple candidates on the epipolar lines and ambiguities in the establishment of correspondences, which can often not be solved. For that reason MATCH-T uses image pyramids as a coarse-to-fine method, thus limiting the length of the epipolar lines at each level of the pyramid. Furthermore, the 'general overkill philosophy' of MATCH-T (Ackermann, 1994) solves some of the problems connected with DEM data acquisition algorithmicly: By switching from interpolation to adjustment, outliers generated by false matches or by objects above the terrain surface (like single trees) can be removed by robust surface fitting, and it has been shown that even breaklines can often be detected automatically in the dense datasets of typically more than 500'000 surface points per stereo pair.

In the method presented in this paper, the length of the epipolar lines is basically unrestricted, and ambiguities due to multiple candidates on the epipolar lines are solved by epipolar line intersection techniques in multiple images, thus avoiding any smoothing effects introduced by surface fitting techniques or by patch sizes in area based techniques.

EPIPOLAR LINE INTERSECTION

The concept of the epipolar line for the establishment of image correspondences is well known in photogrammetry: If the orientation and camera calibration parameters of images are known, for each point P' in one image an epipolar line in an other image can be defined on which the corresponding point P" has to be found. The length of this line can be restricted if approximate knowledge about the depth range in object space is available. Adding a certain tolerance width to this epipolar line segment (due to data quality and measurement errors) the search area for the corresponding point location in the other image becomes a narrow two-dimensional window. Depending on the number of detected points per image, the arrangement of points and the complexity and depth extension of the object a problem of ambiguities will occur here, as often more than one candidate will be found in the search area. There is a general trade-off in stereo matching between the requirements of precision (requiring long epipolar lines and thus also long baselines) and the requirements of reliability (requiring short epipolar lines and thus short baselines to reduce the number of ambiguities). If additional point features do not allow for a reliable distinction of candidates, these ambiguities can often not be solved by a system based on only two cameras. It has been shown in that the probability of such ambiguities grows linearly with the depth extension in object space, linearly with the width of the search area and with the square of the number of points (Maas, 1992a). A solution of this problem based on the intersection of epipolar lines in an image triplet has been shown in (Maas, 1991), where trinocular correspondences in digital images of some 1000 particles marking a turbulent flow were established. The same technique has been applied in combination with a projected dot raster for surface determination in (Maas, 1992b). Also an extension to a four-camera system is shown in (Maas, 1992a). It can be shown that a reasonably chosen additional camera station reduces the probability of ambiguities by a factor of 5 to 100 - depending on the number of points, object geometry and data quality.

The method has meanwhile been extended from trinocular vision to n-ocular vision to be used with an arbitrary number of image coordinate datasets, where the data (= image coordinates of discrete points) may origin form digital, analog or hybrid systems in close range applications as well as in aerial photogrammetry. The aims of this development were:

- The method should be applicable for the establishment of image correspondences between image coordinate datasets from an arbitrary number of images (with the additional requirement that the images have been taken from at least three different camera stations).
- Features which would allow for a distinction of points are not needed (but can optionally be employed).
- Only minimum knowledge about the object space is required. This knowledge can be reduced to rough approximations of the volume boundaries in object space; approximate values for the features are not required, the features may be arbitrarily distributed over the object space. A continuous object surface is not required, or the object surface may show an arbitrary complexity; this includes objects freely distributed in space (like particles in water), occlusions or targets fixed on wires.
- The method is robust against missing points: Not all points have to be measured in all images, points may be missing due to occlusions, failures in the detection procedure, illumination effects, limitation of the sensor format, etc.

The tool was for example used for deformation measurements on a masonry wall using 18 images (Dold/Maas, 1994) and for the modeling of a human head from 40 images (Ursem, 1994).

The aims of the development were mainly derived from tasks occuring in close range photogrammetry, and they exceed the problems occuring in automatic DEM generation, especially concerning the allowance of arbitrary discontinuities in the object. Nevertheless, the application of the technique to the generation of DEMs promises some advantages over conventional methods, justifying research on the benefits of a consequent exploitation of the strength of multi-station geometry in aerial applications.

As the computational effort connected with the strict solution of ambiguities as published in (Maas, 1992a) grows exponentially with the number of images, a reduced version had to be implemented. While the strict solution with three or four images is based on a combinatorics algorithm, the reduced solution is based on probability measures for potential correspondences, which are defined by the number of images the candidate can be traced through. This is implemented in a recursive manner in a way that the longest traces (i.e. the points which can be matched successfully in the largest number of images) are accepted first and in case of ambiguities only a candidate with a trace, which is significantly longer than the traces of the other candidates, is accepted. The flow chart scheme in Figure 1 elucidates this principle.

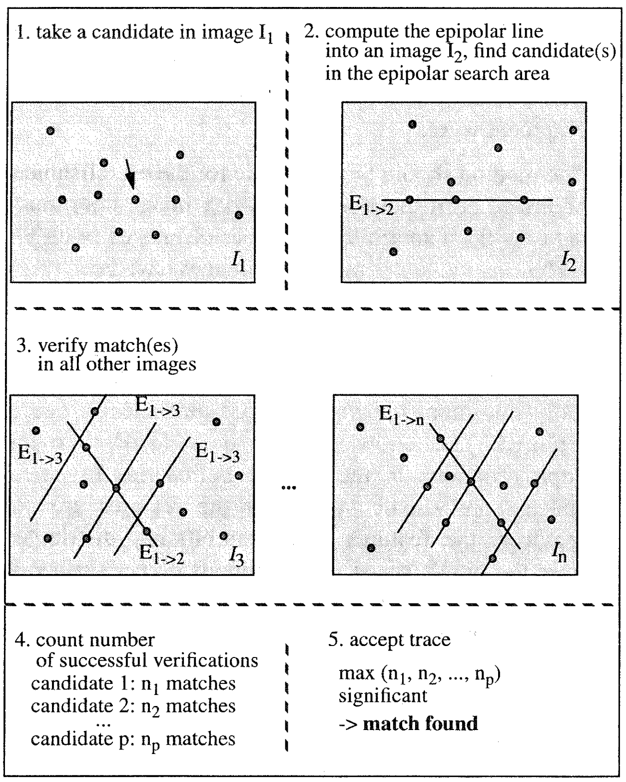


Figure 1: Computation scheme for the automatic establishment of correspondences via epipolar line intersection

For a point P' in the first image the epipolar line in the second image is computed. Then, for all candidates on this epipolar line, the assumption of a correct match is verified or neglected by intersections with the epipolar line of P' in all other images. Finally, the number of successful verifications is counted and the match is accepted if the number of successful matches for a candidate is large enough and significantly larger than the number for all other candidates. This procedure does not require all points to be imaged or detected in all images. Moreover, traces do not necessarily have to begin in the first image. In this manner, all candidates in the epipolar search area are either confirmed as valid candidates or rejected as false or spurious matches.

The technique does not necessarily require base components forming a triangle. Using a parametrization on the epipolar line, it can also be applied to image data acquired from multiple camera positions on a straight line, i.e. with parallel base components and parallel epipolar lines (Maas, 1992a).

A similar approach is also used by (Lotz/Froeschle, 1990) and a number of other authors, who combine short and long base components to warrant both reliability and precision by reducing the probability of the occurrence of ambiguities in the establishment of correspondences between two images. However, this approach does not try to use the multi-image geometry to solve occuring ambiguities. It can however be shown (Maas, 1992a) that if occuring ambiguities are solved the ideal baselength ratio between three collinear camera stations for achieving a maximum rate of successfully established matches is $b_{12} : b_{23} : b_{13} = 1 : 1 : 2$.

A procedure for the solution of ambiguities has also been published under the title multiple-baseline stereo by (Okutomi/Kanade, 1993), who treat the trade-off between precision and reliability in matching with an interesting closed solution avoiding search procedures, which is however limited to one-dimensional area-based matching by minimizing the sum of squared differences.

PRACTICAL RESULTS

Besides to a number of applications in close-range photogrammetry, the n-ocular extension of the epipolar line intersection technique has been applied to the derivation of digital elevation models on two datasets:

- Dataset 1 consists of six scanned aerial images of the Simplon pass area in Switzerland with 80%/60% overlap.
- Dataset 2 consists of a block of 50 images taken from a helicopter with a high resolution digital stillvideo camera.

Dataset 1 - scanned aerial images

Images of a region on the Simplon pass in the Swiss Alps, which shows elevations between 1400m and 2400m, were taken from a flying height of 4600m, using a 150mm lens. The nominal overlap in flight direction was 80% and the overlap between stripes 60%, with relatively large deviations due to terrain height changes. The black-and-white images were scanned at 600dpi resolution (42 micron pixel spacing, image size ~ 5200 x 5200 pixel) on an uncalibrated Agfa Horizon desktop scanner. The fiducial marks in the scanned images were measured with least-squares-matching. A σ_0 of the affine transformation for inner orientation of 60-70 micron indicates the limited accuracy of the scanner.

For the DEM generation with the epipolar line intersec-

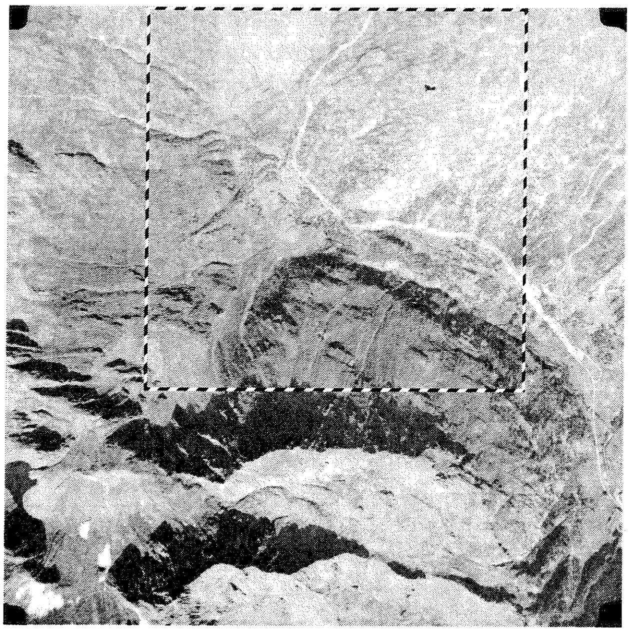


Figure 2: Simplon area scanned aerial image, common region marked

tion technique a sub-block of 2 x 3 images was taken out of the project, with approximately 36% of the image format covering a common region. The camera orientation data was taken from a former aerial triangulation of the whole block on an analytical plotter.

Discrete points were extracted from the six images of the sub-block by an interest-operator (Foerstner, 1986).

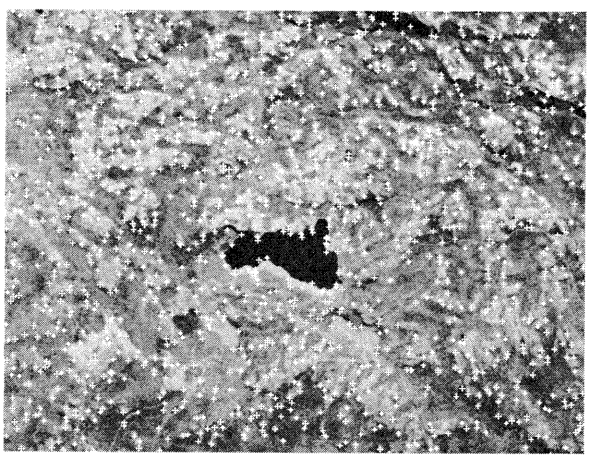


Figure 3: Subregion with interest-operator extracted points

With the Foerstner-operator between 50'000 and 80'000 image points were extracted in the overlapping regions of the six images, which were slightly enlarged to 45% of the whole image area to allow for slight deviations from the flight plan and for effects of the topography on the size of the image overlap. These points were fed into the epipolar line intersection routine, which was able to match a total of about 1500 points in all images. The original result without any post-processing is shown in Figure 3. Note that no post-processing has been applied to these data. With only about 1500 points matched in all 6 images the yield is relatively sparse and not quite satisfactory. On the other hand, however, the blunder rate is extremely low as compared to other automatic DEM generation techniques: In Figure 4 only 5-6 blunders can be detected, which corresponds to a blunder rate of only 0.3%. Moreover, the blunders show up as clear peaks or holes in the visualization can easily be removed by local post-processing methods like median filtering or robust surface fitting.

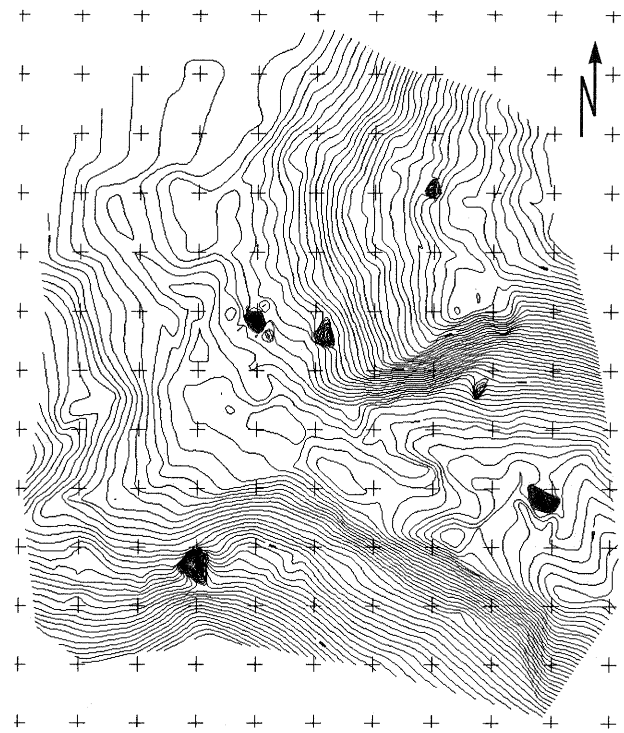
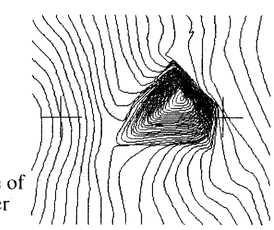


Figure 4: Simplon test area, 10m contour lines (orientation and scale do not match with Figure 2)



When in addition to the points which were detected in all six images also those points were accepted which were matched in any 5 of the 6 images, 5500 points could be matched, but with a larger percentage of blunders of approximately 2%. The relatively sparse result can be explained by the following two reasons: One problem in the data processing was the poor geometric stability of the uncalibrated scanner, which necessitated to work with a rather large tolerance to the epipolar lines (80 micron in this case) to make sure that corresponding points were not rejected due to inaccuracies in image space. As the number of ambiguities in the establishment of correspondences grows with the square of the tolerance to the epipolar line, this fact leads to an increasing number of correct matches being rejected as spurious matches due to unsolvable ambiguities. Another reason is the probability of the interest operator detecting identical points in all involved images, which decreases exponentially with the number of images.

Dataset 2 - stillvideo images

For a pilot study on the applicability of high resolution digital stillvideo cameras in aerial photogrammetry a block of 5 x 10 images was taken with a Kodak DCS200 (8 bit monochrome 1524 x 1012 pixel sensor, format 14mm x 9mm, pixel spacing 9µm x 9µm, internal harddisk for 50 uncompressed images, 18mm lens) from a helicopter over a village in Grisons. The average image scale was 1 : 18'000, the average overlap was 70% in both directions, with large deviations due to handling difficulties with the hand-held camera in the unsteady helicopter. The orientation data of the images was taken from a self-calibrating bundle block adjustment after semi-automatic measurement of the image coordinates of 36 signalized points. Checkpoint accuracies of 2cm for the planimetry coordinates and 4cm for the height coordinate were achieved in this aerotriangulation (Kersten, 1996).

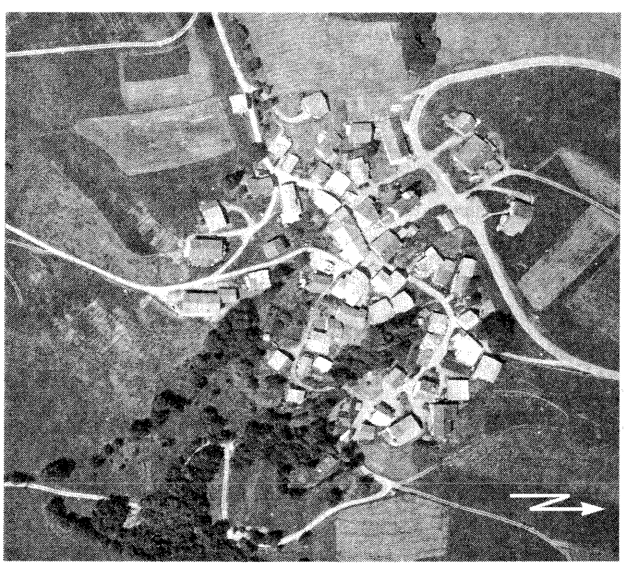


Figure 5: DCS200 aerial overview image of a village

An average of 7000 points per image was extracted by the Foerstner-operator. Due to the geometric quality of the data a maximum tolerance of 10 micron (~ 1 pixel) to the epipolar line was found to be appropriate for the acceptance of possible matches. Applying the epipolar line intersection routine with the requirement that each point has to be matched in at least 5 images, a total of 6860 object points could be reconstructed. The maximum number of images a single points was successfully matched in was 21. The average standard deviation of the reconstructed points was 0.027/0.027/0.109m in X/Y/Z. An contour line plot generated from the 6860 object points is shown in Figure 6. Again, no post processing was applied to the raw data provided by the routine. The plot looks quite good in the surroundings of the village, but rather chaotic within the village.

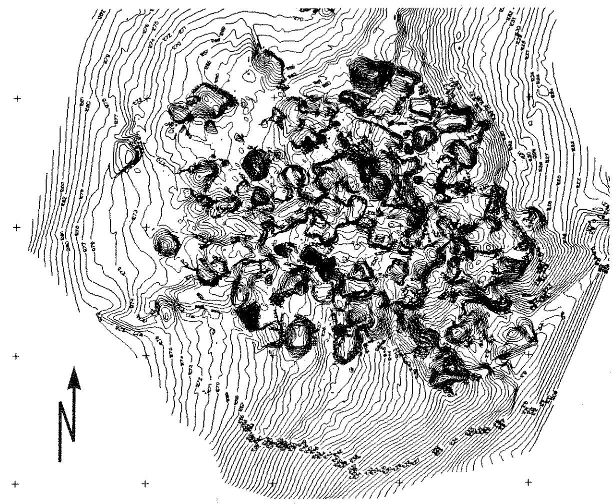
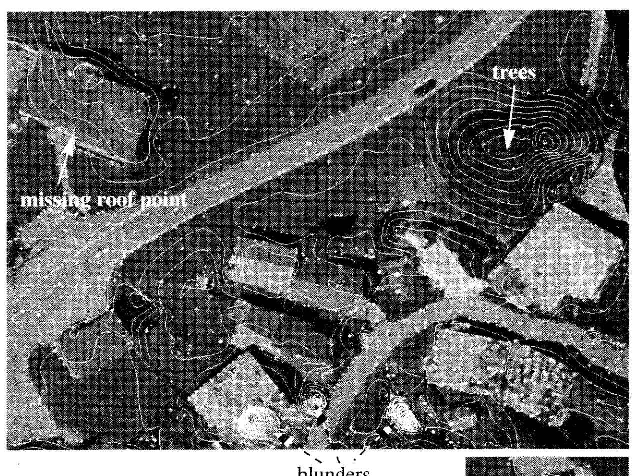


Figure 6: Urmein test area, 1m contour lines (orientation and scale do not correspond to Figure 5)

Some problematic regions inside the village are shown in parts of a digital orthophoto mosaic with contour line overlay in Figure 7: As no information on buildings, trees, etc. was a priori given or explicitly extracted by the technique, a large number of successfully matched points are not situated on the terrain surface, but on objects above the surface. Due to the characteristics of the interest operator and the loss of points caused by the probability of the interest operator detecting identical points in all involved images, however, buildings are often modeled only rather incompletely. In some cases only one or two points of a roof were matched, and terrain points at the foot of buildings are often missing.



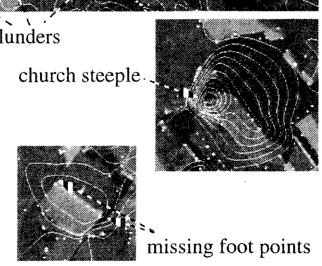


Figure 7: Insuffently modeled parts (orthophoto with matched points and 2.5m contour line overlay)

A thorough visual check of the digital orthophoto mosaic with an overlay of matched points and contour lines showed a total of only 12 gross errors in the dataset, which corresponds to an error rate of 0.2%. Sharpening the requirement for the acceptance of points to matches in at least 6 images, 4270 object points with 5 gross errors remained, and if points were only accepted when matched in at least 7 images, the number of reconstructed object points drops to 2750 (see Table 1).

Table 1: Interest operator efficiency

min. images	5	6	7	8	9	10
object points	6866	4272	2757	1787	1187	757

In combination with the average overlap of ~70% (theoretical average visibility of points in 12.25 images) this indicates a probability of approximately 65% for the Foerstner-operator detecting identical points in two images and a probability of approximately $0.65^{(n-1)}$ for detecting identical points in n images.

CONCLUSION

In contrast to most stereo-based techniques, the presented technique does not avoid ambiguities in the establishment of correspondences by image pyramids, but solve them by a consequent exploitation of the geometric strength of multiple images, implemented via the intersection of epipolar lines or multi-baseline techniques. The method shown in the presentation can be characterized by the following keywords:

- High reliability due to the multi-image geometry.
- No approximate values are needed, no interactively set start points, no image pyramids are involved.
- No a priori knowledge on maximum terrain slopes is required, the terrain slope can be practically unlimited.
- Matching of discrete points avoids smoothing effects connected with many area-based DTM generation methods.
- Good precision due to subpixel interest-operator and measurement in multiple images.

As far as the allowable arbitrary terrain discontinuities are concerned, these options exceed the requirements posed by the automatic generation of digital elevation models in most terrain types. However, the aim of the study to show the potential of the consequent use of the strength of multi-image geometry for the robust establishment of correspondences, could be achieved. Blunder rates of less than 0.5% of the matches without the use of pyramid approaches or given approximate values can be considered a very good result for automatic DEM generation in mountain regions.

One limitation of the technique is the requirement that points have to be imaged in multiple images, which necessitates 80%/60% image overlap in mountain regions. A major limitation is the probability of interest operators detecting identical points in multiple images. In the image material used in the presented study, the probability that a point, which was detected by the Foerstner operator in one image, was also detected in a second image, was found to be approximately 65%, reducing the number of successful matches exponentially with the required number of images. Although possibly not dense enough for a good terrain description, the presented results may be very useful as approximate values for other techniques.

References

- Ackermann, F., 1994: Digital elevation models - techniques and application, quality standards, development. IAPRS Vol. 30, Part IV, pp. 421-432
- Baltsavias, M., Stallmann, D., 1992: Advancement in matching of SPOT images by integration of sensor geometry and treatment of radiometric differences. IAPRS Vol. XXIX, Part B4, pp. 916-924
- Dold, J., Maas, H.-G., 1994: An application of epipolar line intersection in a hybrid close range photogrammetric system. IAPRS Vol. 30, Part V
- Förstner, W., 1986: A Feature Based Correspondence Algorithm for Image Matching. IAPRS, Vol. 26, Part 3/3, pp. 150-166
- Grün, A., 1985: Adaptive least squares correlation: A powerful image matching technique. South African Journal of Photogrammetry, Remote Sensing and Cartography 14 (3), pp. 175-187
- Kersten, Th., 1996: Aerotriangulation mit einer digitalen Still Video Kamera Kodak DCS200. VPK 2/96
- Krzystek, P., 1991: Fully automatic measurement of digital terrain models. Proc. of the 43. Photogrammetric Week, Stuttgart, Germany, pp. 203-214
- Lotz, R., Fröschle, E., 1990: 3D-Vision mittels Stereobildauswertung bei Videobildraten. 12. DAGM-Symposium Mustererkennung, Informatik Fachberichte 254, Springer Verlag
- Maas, H.-G., 1991: Digital Photogrammetry for Determination of Tracer Particle Coordinates in Turbulent Flow Research. Photogrammetric Engineering & Remote Sensing Vol. 57, No. 12, pp. 1593-1597
- Maas, H.-G., 1992a: Complexity analysis for the determination of image correspondences in dense spatial target fields. IAPRS, Vol. XXIX, Part B5, pp. 482-485
- Maas, H.-G., 1992b: Robust Automatic Surface Reconstruction with Structured Light. IAPRS, Vol. XXIX, Part B5, pp. 102-107
- Okutomi, M., Knade, T., 1993: A Multiple-Baseline Stereo. IEEE transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No. 4, pp. 353-362
- Schewe, H., 1987: Automatic Photogrammetric Car-Body Measurement. Proc. of the 41. Photogrammetric Week, Stuttgart, Germany, pp. 47-57
- Ursem, R., 1994: Accurate Reconstruction of a 3D Wireframe of a Human Head. Diploma Thesis TU Delft, The Netherlands