Y VINVESTIGATION OF GEOMETRIC CONSTRAINTS FOR MATCHING HIGH RESOLUTION SATELLITE IMAGES

P. M. Dare

Airborne Research Australia, Flinders University, PO Box 335, Salisbury South, SA 5106, AUSTRALIA
Paul.Dare@AirborneResearch.com.au

Commission III, WG II/2

KEY WORDS: Ikonos; DEM/DTM; Matching; High resolution; Satellite imagery

ABSTRACT

The creation of digital surface models from stereo imagery is a well-understood procedure that is central to digital photogrammetric processing. Recently, however, attention has focused on the creation of surface models from high resolution satellite imagery, which is not quite so straightforward due to the specific attributes of spaceborne imaging systems, and the fact that some data suppliers do not release details of the sensor and camera models. This paper describes a matching procedure for creating digital surface models (DSMs) from stereo imagery acquired by the high resolution Ikonos satellite. Central to this matching procedure are the geometric constraints that are commonly used to reduce the search space and hence constrain the matching solution. Results are presented of the use of two different geometric constraints (one image space constraint, and one object space constraint) applied to two very different Ikonos stereopairs. A range of digital surface models were created, which, when compared to reference data, showed height differences of less than a few metres. Furthermore, visual evaluation of the resulting surface models showed that both geometric constraints yielded a good representation of the true surface.

1. INTRODUCTION

The topic of automatic image matching, especially for the purpose of surface modelling, has received considerable attention for many years. (For a concise historical summary, the reader is referred to Samadzadegan, 2002.) Although many early problems associated with image matching have been resolved by the development of new algorithms, or the application of high-end technology, new issues continue to arise. The majority of previous image matching research, from a geometric point of view at least, has utilised aerial photography or moderate resolution satellite imagery (such as SPOT panchromatic data). Nowadays the wide availability of high resolution stereo satellite imagery means that image matching can be thoroughly investigated. The different attributes of these sensors, as compared to aerial photography, mean that new problems have to be resolved, and hence new algorithms or matching strategies have to be developed.

Matching conjugate points in high resolution stereo satellite imagery is more challenging than in air photos due to the far more limited opportunities for satellite image acquisition. Aerial photography that is recorded for the purpose of topographic mapping would always be acquired under ideal conditions, namely good illumination, appropriate base to height ratio for the level of terrain undulation, and correct scale for the ground features being imaged. With high resolution satellite imaging these parameters can rarely be changed. Illumination is dependent upon season and latitude (time of day is fixed by orbital parameters); base to height ratio is set by the satellite operator; and, scale is fixed by sensor resolution, orbital height and look angle.

The two main consequences of the difference between automatic image matching with high resolution satellite imaging, and with aerial photography, are firstly that alternative matching strategies may be required to account for the lack of sensor orientation information, and secondly that results cannot be expected to be as good as those from aerial photography (a result confirmed by Fraser et al., 2001; Fraser et al., 2002a).

This paper presents the results of a study which investigated different matching strategies for pairs of Ikonos images. In particular, the study has focussed on methods used to constrain the search space. As a result, two different geometric constraints have been evaluated. The first constraint, based on epipolar geometry, operates in image space, and can be applied to all images, whether or not they are aligned to epipolar coordinates. The second constraint is based on the affine projective model (Fraser et al., 2002b), which like rational polynomial coefficients (RPCs), maps image space coordinates to object space coordinates. Matching is constrained by limiting the search for conjugate points along vertical nadir lines in the stereomodel. By applying these constraints to two Ikonos stereopairs, the relative advantages and disadvantages of each constraint have been compared.

2. THE AFFINE PROJECTIVE MODEL

The affine projective model is somewhat similar to the RPC model in that it relates image space coordinates to object space coordinates without any knowledge of the sensor model or exterior orientation (EO). The general form of the model describing an affine transformation from 3D object space (*X,Y,Z*) to 2D image space (*x, y*) for a given point *i* is expressed as:

$$\begin{aligned} x_i &= A_1 X_i + A_2 Y_i + A_3 Z_i + A_4 \\ y_i &= A_5 X_i + A_6 Y_i + A_7 Z_i + A_8 \end{aligned} \qquad (1)$$

This model comprises eight parameters per image, these accounting for translation, rotation, and non-uniform scaling and skew distortion. Implicit in (1) are two projections, one scaled-orthogonal and the other skew-parallel. With high resolution satellite imaging systems such as Ikonos and Quickbird with narrow fields of view, the assumption that the projection is parallel rather than perspective has been shown in practical tests to be sufficiently valid (Fraser et al., 2002b). In the reported implementation of the affine projective model, all model parameters are recovered simultaneously along with triangulated ground point coordinates in a process analogous to photogrammetric bundle adjustment.

3. DATA USED IN THIS STUDY

Two pairs of satellite images were used in this study. Firstly a stereo pair of Ikonos images, resampled according to epipolar geometry by the data supplier and covering a 7 x 7 km area over the city of Melbourne, Australia, was selected. The terrain does not vary significantly in the area, with the lowest point at sea level and the highest point at about 50m above sea level. The Central Business District is located at the centre of the area, and contains buildings up to 250m tall. About 15% of the imagery depicts water at sea level. One of the images from the stereopair is shown in figure 1.

Figure 1. Ikonos image of Melbourne

The second image pair used in this study was a non-epipolar Ikonos stereopair of San Diego, USA. Although this image pair covered a wide ground area, a much smaller area of approximately 7km x 5km was extracted. This sub-sampled region was chosen in order to provide a very different test area to the Melbourne data. Consequently it features mountainous terrain, vegetation land cover and no urbanisation. One image of this stereopair is shown in figure 2.

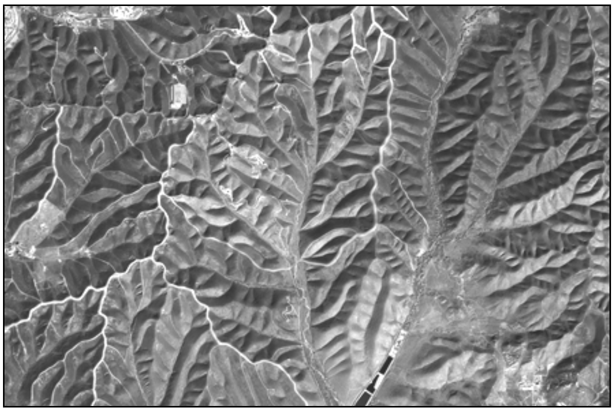

Figure 2. Ikonos image of mountainous region near San Diego

For each data set, the parameters of the affine projective models were calculated using GPS-surveyed ground control points. Most of the ground control points observed were road roundabout centres, easily measured in image space by taking the centroids of ellipses fitted to multiple edge points around each roundabout. The remaining control points were road and building corners and other distinct features conducive to high precision measurement in both object space and image space.

4. GEOMETRICALLY CONSTRAINED IMAGE MATCHING

4.1 Background

In any image matching process there are generally three key steps: selection of candidate points; definition of search space; and, comparison of similarity measures. The most important, in terms of practical implementation, is the definition of the search space. By choosing the appropriate search space, computation time is kept to a minimum and the potential for finding blunders is reduced. In typical image space matching the search space is a two dimensional area centred on the pixel being matched. The search area has to be large enough to ensure the correct match can be found, but not too large that the processing time becomes computationally absurd. Thus, any way of reducing the search space, but retaining the guarantee of the existence of a correct match, is a significant improvement to any matching algorithm. This is what geometric constraints aim to achieve.

4.2 Epipolar constraint

The most common geometric constraint used in image matching is the epipolar constraint, which allows the search space to be reduced from a two dimensional area to a one dimensional line. By reducing the search space in this way, the speed of matching algorithms can be increased by an order of magnitude, and the chances of finding blunders is greatly reduced. With aerial photography, epipolar lines in stereo images are usually determined from a knowledge of the EO parameters. With high resolution satellite imaging the EO is generally unknown, meaning that the imagery must be purchased in its normalized (i.e. epipolar projected) form.

Matching points in a stereopair of images that are aligned to epipolar geometry is a simple task since it only involves a one dimensional search space. However, by taking advantage of the epipolar geometry it is assumed that the alignment to epipolar geometry is error-free. Since the normalization of the Ikonos imagery has been carried out by the supplier of the data (Space

Imaging Inc.) the quality of the normalization process is unknown.

It is, however, simple to perform a check of the epipolarity of the stereo images. for the purposes of this study, groups of conjugate points located at the extremities of the image were matched using a geometrically unconstrained matching algorithm. The matched points were then filtered to remove all points with a normalized cross correlation coefficient of less than 0.9. By comparing the y values (line numbers) of the remaining well-matched points, it was possible to assess the quality of the algorithm that has created the epipolar images. The results of the matching of the four groups of points are shown in table 1.

Group of points	Number of well-matched points	RMS y residual (pixels)	Mean y residual (pixels)
Top left of image	149	0.54	-0.40
Top right of image	69	0.57	-0.44
Bottom left of image	109	0.65	-0.51
Bottom right of image	82	0.73	-0.44

Table 1. Results of matching at extremities of the image pair.

The results presented in table 1 clearly show that the alignment of the Melbourne stereopair to epipolar geometry is not exact: there is apparently a systematic shift in the y direction of approximately half a pixel between the two images. Thus, throughout the rest of this study a more loosely defined epipolar constraint has been used where the search space is limited not to one single line, but to a two dimensional area extending to one line on either side of the epipolar line. Since the matching process is more loosely constrained, the processing time necessarily increases.

4.3 Alignment of non-epipolar images to epipolar geometry

Since not all available Ikonos data is aligned to epipolar coordinates (as is the case with the San Diego data used in this study) the re-alignment of non-epipolar images to epipolar geometry was investigated. There are various methods available for achieving this, with some being more rigorous than others. Three possible methods include using the satellite ephemeris data (or strictly speaking, the image metadata), the fundamental matrix, or a second order polynomial model.

Although the data supplier does not release detailed satellite orientation data, a small amount of limited image metadata is supplied with images. This data includes details such as the time of acquisition, the sun azimuth and elevation, the approximate scene location, but most importantly (to this study at least) is the satellite azimuth and elevation. Using these approximate values is it possible to estimate the angles of rotation required to align the stereopair to epipolar coordinates. However, the estimate is coarse and of unknown accuracy. Additionally, it does not take into account that the images are time-dependent linescanner images, and not frame photographs.

An alternative, and much more rigorous approach, is to implicitly determine the relative geometry of the images and hence derive the fundamental matrix, thus allowing a direct mapping of points between the images (Luong and Faugeras, 1996). Although this technique of determination of epipolar geometry is more commonly associated with computer vision than photogrammetry, there is no reason why it cannot be applied to Ikonos images. Unfortunately problems are not uncommon during the process of deriving the fundamental matrix, since instabilities in the solution can occur. Once again, this method does not account for the linescanner nature of the data.

A third method of approximating the epipolar geometry between images, described by Zhang et al. (2002), is based on a second order polynomial model proposed by Orun and Natarajan (1994). The model assumes that during image acquisition the pitch and roll angles remain constant, but the yaw angle variation follows a second order polynomial. The result is an epipolar curve, rather than an epipolar line, defined by the quadratic polynomial:

$$\begin{aligned} y_r &= (a_0 x_l + a_1 y_l + a_2) + (a_3 x_l + a_4 y_l + a_5) x_r \\ &+ (a_6 x_l + a_7 y_l + a_8) x_r^2 \end{aligned} \tag{2}$$

where (x_l, y_l) and (x_r, y_r) are conjugate points in the left and right images respectively, and a_0 to a_8 are parameters to be determined. As can be seen, each unique point in the left image can be associated with a unique epipolar curve in the right image: knowledge of the coordinates of a point in the left image allows an epipolar curve to be plotted, upon which the conjugate point will lie. The parameters a_0 to a_8 have to be determined in advance, which is done by substituting the coordinates of known conjugate points into equation (2) and solving as required.

In order to validate the quality of the second order polynomial epipolar model, tests were carried out with the Melbourne stereo data, which are known to be already aligned to epipolar geometry. Initially 648 points were matched between the two images using a non-constrained hierarchical matching strategy. Of the 648 matched points, 108 were found to have cross-correlation coefficients of greater than 0.95, implying well matched points. These 108 well matched points were then split into two groups, one of which (54 points) was used to calculate the parameters (a_0 to a_8) of the second order polynomial model. Although only nine points are required to obtain the solution directly, 54 points provided a good deal of redundancy, and allowed the parameters to be estimated by a least squares method. The remaining 54 well matched points were used as check points to verify the accuracy of the second order polynomial model. By substituting the coordinates of the matched point from the left image (x_l, y_l) into the model, along with the abscissa of the matched point from the right image (x_r), a new value of y_r was calculated and compared to the expected value (from the coordinates of the matched point in the right image). This process was repeated for all 54 check points. Consequently it was found that the root mean square discrepancy was marginally less than one pixel, implying that the second order polynomial epipolar model is sufficiently accurate to be used as a geometric constraint for matching non-epipolar stereo images.

Confident that the Melbourne images could be successfully aligned to epipolar coordinates, the San Diego images were similarly processed. As with the Melbourne data, well matched points (matched points with cross-correlation coefficients of greater than 0.95) were used to determine the parameters a_0 to a_8. This time a total of 34 points were used in the least squares estimation process.

4.4 Affine projective model constraint

The affine projective model constraint is an object space constraint. Rather than limiting the search to lines in image space, the search is limited to lines in object space. This is done by generating a grid of object space points (*X, Y, Z*) which coincides with the coverage of the stereomodel. These points are sequentially projected into the image space of both images being matched, using the appropriate affine model. The image points are then matched and the similarity measure recorded. The process is then repeated with a new value of *Z* for the object space point. The value of *Z* (basically the height component of the point in object space) is varied until a highly correlated match is found. At this point, the value of *Z* will be a good estimation of the height of the terrain.

5. MATCHING STRATEGY

5.1 Candidate matching point selection

Selection of candidate matching points was carried out automatically. A regular grid of points was projected across the images being matched. Since each pair of images in each data set were georeferenced to ground coordinates (i.e. pre-aligned with each other), the projected grids of points could be considered as rough approximations of conjugate points, and hence used in the first stage of the matching algorithm. This method of initial point selection was found to be both efficient and sufficiently accurate for the matching strategies under investigation.

5.2 Matching strategy

Throughout this study, the matching strategy used for determining the conjugate points incorporates intensity-based matching and utilises the cross-correlation coefficient as a similarity measure. Since this is the most appropriate strategy for matching points in images with similar radiometric distributions, such as stereopairs of high resolution satellite imagery where time difference between image acquisition is small, no other matching methodologies (such as least squares matching or feature-based matching) were tested.

For each point to be matched, an image chip from the master image is repeatedly projected into the slave image at various locations around the likely candidate match point (the spatial extent and frequency of these locations being defined by the search space and step size respectively). The location where the maximum value of the cross-correlation coefficient, γ, is found indicates a successful match. The cross-correlation coefficient, γ, is given by:

$$\gamma(x, y) = \frac{\sigma_{MS}}{\sigma_M \sigma_S} \tag{3}$$

where σ_M and σ_S are the standard deviations of the master and slave chips being matched and σ_{MS} is the covariance of the intersection of the master chip with the slave chip (Gonzalez and Woods, 1992).

5.3 Hierarchical matching

Hierarchical matching is a technique often used in image matching in order to reduce processing time. It has been incorporated into the matching strategy used in this study by repeating the matching process whilst progressively reducing the search space and the step size. In the first iteration a wide search space is used with a large step size. The result of the first iteration (a coarse match) is used as the approximation of the next iteration, where both the search space and the step size are decreased. The matching is then repeated, and the result (this time less coarse) is again passed to the next iteration. Thus in each iteration the matching results become progressively more refined, but processing time remains reasonable. The number of iterations used in the matching in this study was four.

6. RESULTS

6.1 Analysis of correlation coefficients

Using the methodology described above with the two geometric constraints (epipolar and affine), grids of points were matched in each of the test images. The matching results for these experiments are presented in table 2.

Test site	Geometric constraint	No. of points in grid	No. of matched points with correlation coefficient > 0·8
Melbourne (epipolar)	Epipolar	58302	37473 (64·3%)
	Affine	68199	43637 (64.0%)
San Diego (non-epipolar)	Epipolar	17673	6240 (35.3%)
	Affine	18149	5613 (30.1%)

Table 2. Matching results

Table 2 shows the number of matched points with correlation coefficients greater than 0.8, and the corresponding percentages with respect to the total number of points matched, for each combination of image and matching constraint. A number of very interesting conclusions can be drawn from these results, since, as it can be seen, the epipolar constraint and the affine constraint give very similar results, but these results are very image dependent.

The first important point to note is that use of the epipolar constraint based on the second order polynomial used with the San Diego data does not adversely affect the matching results. If this model had been incorrect, then the number of well matched points would have been much less than the 35% that was achieved.
The second point to note is the effect of image content on the matching algorithm. The percentages of well matched points for the San Diego data are much lower than those for the Melbourne data. This could be due to a number of reasons. Firstly, the terrain varies much more steeply in the San Diego data, meaning object space features may appear differently in each of the images due to the differing incidence angle. Secondly, the steeply varying terrain causes differing solar reflections, meaning that ground features may have differing radiometric signatures in each image. Finally, and probably most importantly, the content of the San Diego images is very

different to that of the Melbourne images. The San Diego area has a much more homogeneous texture since it is almost entirely forested. Conversely, the Melbourne scene is an urban area with a more heterogeneous texture and perhaps more ideally suited to image matching.

6.2 Comparison with reference data

Although an analysis of the correlation coefficients can be instructive, it does not reveal sufficient information by which to assess the success of the matching algorithm and the quality of the matched points: it is entirely possible that some of the well matched points with correlation coefficients greater than 0.8 may be blunders. Further conclusions can be inferred by triangulating the matched points to give object space coordinates (using the affine model), and comparing those coordinates with a reference digital elevation model (DEM).

This process was carried out with the Melbourne data where a high quality reference DEM was available. This DEM was created from aerial photography and resampled to a grid spacing of 25m. Significant manual post-processing was carried out on this DEM to ensure it gave an accurate representation of terrain heights. As a result, it would be expected that points triangulated from the matched Ikonos points would be uniformly higher, since they represent the surface of the objects in the imagery, and not the terrain.

Geometric constraint	Matched points with correlation coefficient > 0.8	
	Mean (m)	Std. Dev. (m)
Epipolar	3.41	6.80
Affine	3.64	7.60

Table 3. Differences between triangulated points and reference DEM – Melbourne Test Area

As expected, it can be seen from table 3 that the differences between the triangulated points and the reference DEM are all positive, and of a magnitude that could realistically represent the mean building heights within the test areas. Also noteworthy is the fact that the differences between the affine constraint results and the epipolar constraint results are small, although the surface created from the points matched using the affine constraint is marginally higher than the surface created from the epipolar constraint.

Since no reference DEM was available for the San Diego test area, a similar comparison was not possible.

6.3 Visual analysis

Finally a visual analysis of the results was carried out. A surface model was created for each test area from the triangulated points. Figure 3 shows a 5m grid DSM of the Melbourne test area, while figure 4 shows a similar DSM of the San Diego test area. In each figure the lighter tones of grey represent high elevation, whilst the darker tone represent low elevation.

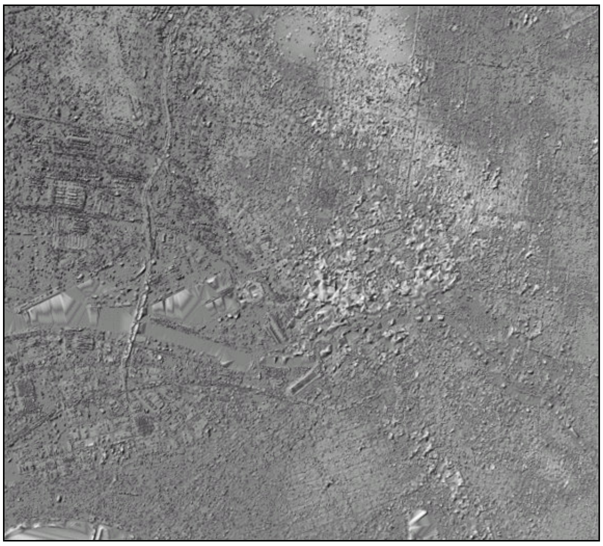

Figure 3. DSM of Melbourne test area

Figure 3 clearly illustrates the potential of matching stereopairs of Ikonos images. The resulting DSM shows many features, including buildings, vegetation, bridges, overpasses and even roads and railways. Although some blunders are apparent, particularly in the river in the eastern part of the test area, the majority of the matching has been successful. Even with the presence of the surface features, it is still possible to recognise the pattern of the underlying terrain, with areas of high ground in the northern part of the test area. Note that this image has been shown from a near-nadir perspective due to the fact the terrain varies very little across the image.

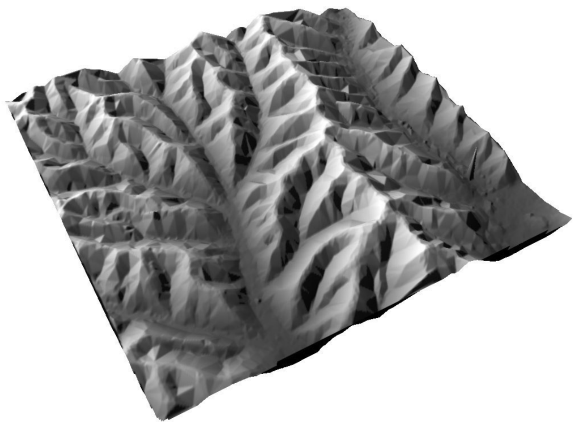

Figure 4. DSM of the San Diego test area

Figure 4 shows the San Diego DSM displayed in perspective view (with the vertical scale exaggerated). It is clear to the observer that the DSM is a very good representation of the underlying terrain, with hydrological and topographic features standing out very clearly. There are very few, if any, obvious blunders.

7. DISCUSSION

This paper has presented the use of two different geometric constraints on two stereopairs of high resolution satellite imagery. The results of the matching procedures have been analysed by assessing the correlation coefficients, triangulating individual points, and performing a visual analysis. Consequently these results show that the object space geometric constraint (based on the affine projective model) is certainly as good as the image space geometric constraint (based on the epipolar model). An advantage of the affine-based constraint is that the matching parameters can be specified in ground coordinates. If approximate terrain heights are known, the matching search space can be specified very accurately indeed, and hence processing time and the number of potential errors can be reduced. (With the image space constraint, the matching parameters are much more arbitrary.) Additionally, these results further confirm the usefulness of the affine projective model in geometric processing of high resolution satellite imagery.

Images not aligned to epipolar coordinates (the San Diego data) have also been successfully matched by use of a quadratic epipolar model. Since the results of this model were as good as those from the affine model, it can safely be assumed that use of the quadratic epipolar model is justified. This is an important result since it means that a high resolution image pair (i.e., not a stereopair) can be matched as successfully as a stereopair.

Success of these empirical matching constraint models is largely contingent on the images being free of scanning non-linearities. Fortunately Ikonos appears to be largely free of such effects. Future work will investigate the possibilities of using multi-temporal, same-sensor images for terrain modelling, as well as multi-temporal, multi- sensor images.

8. ACKNOWLEDGEMENTS

This work was carried out while the author was a research fellow at the Department of Geomatics, University of Melbourne. The author would like to thank Prof. Clive Fraser, Dr. Harry Hanley, Mr. Simon Cronk and Mr. Takeshi Yamakawa for their assistance with various aspects of the research.

9. REFERENCES

Fraser, C. S., Baltsavias E. P., Gruen, A., 2001. Ikonos Geo stereo images: geometric potential and suitability for 3D building reconstruction. In: Photogrammetrie-Fernerkundung-Geoinformation: Geodaten schaffen Verbindungen, E. Seyfert (Ed.), Publications of the German Society for Photogrammetry and Remote Sensing, 10: 113-121.

Fraser, C. S., Baltsavias E. P., Gruen, A., 2002a. Processing of *Ikonos* imagery for submetre 3D positioning and building extraction. *ISPRS Journal of Remote Sensing*, 56(3): 177-194.

Fraser, C. S., Hanley, H. B. and Yamakawa, T., 2002b. Three-dimensional geopositioning accuracy of Ikonos imagery. *Photogrammetric Record*, 17(99): 465-479.

Gonzalez, R. C. and Woods, R. E. (1992). *Digital image processing*. Addison-Wesley, New York, 716 p.

Luong, Q. T. and Faugeras, O., 1996. The fundamental matrix: Theory, algorithms and stability analysis. *International Journal of Computer Vision*, 17(1): 43-76.

Orun, A. B. and Natarajan, K., 1994. A modified bundle adjustment software for SPOT imagery and photography: tradoff. *Photogrammetric Engineering & Remote Sensing*, 60(12): 1431-1437.

Samadzadegan, F., Azizi, A., Lucas, C., Hahn, M., 2002. Automatic Surface Reconstruction Based on the Fusion of Fuzzy Logic and Robust Estimation Techniques. *Photogrammetric Record*, 17(100): 651-674.

Zhang, L., Pateraki, M., Baltsavias, E., 2002. Matching of Ikonos Stereo and Multitemporal GEO Images for DSM Generation. *Proceedings of Map Asia 2002, Asian Conference on GIS, GPS, Aerial Photography and Remote Sensing*, Bangkok, Thailand (on CD-ROM).