View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Ghent University Academic Bibliography

Published as: Tack, F., Goossens, R. & Buyuksalih, G. (2012), Assessment of a Photogrammetric Approach for Urban DSM Extraction from Tri-Stereoscopic Satellite Imagery. *The Photogrammetric Record,* 27(139): 293–310.

ASSESSMENT OF A PHOTOGRAMMETRIC APPROACH FOR URBAN DSM EXTRACTION FROM TRI-STEREOSCOPIC SATELLITE IMAGERY

By FREDERIK TACK (f.tack@ugent.be), RUDI GOOSSENS (rudi.goossens@ugent.be),
Dept. of Geography, Ghent University, Belgium

and GURCAN BUYUKSALIH (gbuyuksalih@yahoo.com)
IMP-Bimtas, Istanbul, Turkey

Abstract

Built-up environments are experienced as extremely complex for 3D surface modelling purposes. Main distortions that hamper 3D reconstruction from 2D imagery are image dissimilarities, concealed areas, shadows, height discontinuities and discrepancies between smooth terrain and man-made features. A methodology is proposed to improve automatic photogrammetric extraction of an urban surface model from High Resolution Satellite Imagery (HRSI) with the emphasis on strategies to reduce the effects of cited distortions and to make image matching more robust. Instead of a 'standard' stereoscopic approach, a Digital Surface Model (DSM) is derived from tri-stereoscopic satellite imagery based on an extensive multi-image matching strategy that fully benefits of the geometric and radiometric information enclosed in the three images. The bundled triplet consists of an IKONOS along-track

pair and an additional near-Nadir IKONOS image. For the tri-stereoscopic study, a densely built-up area extending from the centre of Istanbul to the urban fringe is selected, covering approximately 60 km^2. The accuracy of the model extracted from the IKONOS triplet as well as the model extracted from only the along-track stereopair are assessed by comparison with 3D check points and 3D building vector data.

KEYWORDS: multi-image matching, DSM generation, tri-stereoscopy, IKONOS, urban

INTRODUCTION

STUDIES of the urban environment and its dynamics demand high quality 3D surface models for a wide range of applications. As humans are naturally able to see in three dimensions, a 3D representation of the terrain surface assists the geospatial analysis and understanding of various processes. 3D descriptions of an urban environment and its topography can be generated by various acquisition methods, such as photogrammetry, airborne laserscanning or human-based interpretation and digitalisation to name a few of them.

3D city model generation by highly automated photogrammetric methods, based on large-scale imagery, has great potential but still urban scene and especially building modelling is considered as an extremely hard problem. Research activities at many institutions were or are devoted to the problem of reconstructing a complex 3D environment based on 2D imagery, such as work reported in Ackermann (1983), Gruen (1985), Lemmens (1988), Dhond and Aggarwal (1989), Baltsavias (1991), Vosselman and Haala (1992), Heipke (1996), Krzystek et al. (1996), Julien (1999), DPCOR (2000), Brown et al. (2003) and Zhang and Fraser (2008). The authors wish to refer to Berthod et al. (1995), Weidner and Förstner (1995), Gabet et al. (1997), Förstner (1999), Baillard and Maître (1999), Guo and Yasuoka (2002), Zhang (2005), Jacobsen (2006), Crespi et al. (2007), Lafarge et al. (2008), Alobeid et al. (2010), Kraus and Reinartz (2010) and Woo and Park (2011), where the urban issue and surface/building reconstruction is specifically addressed and reported thoroughly. However, in spite of the many efforts in the photogrammetric and computer vision fields, fully automated surface and building reconstruction from large-scale imagery remains an unsolved problem (Zhang, 2005).

In this paper, a methodology will be proposed to improve the quality of a DSM extracted from HRSI with the emphasis on strategies, applied at different processing levels, to cope with distortions that hamper urban terrain and 3D building reconstruction. The most prominent distortions are shadow, occlusion, image dissimilarities as a consequence of relief displacement and steep changes in slope. Instead of a 'standard' stereoscopic approach, a DSM is generated from an IKONOS triplet based on an innovative multi-image, multi-primitive matching strategy, guided from the object space, which fully benefits of the radiometric, geometric and projective information enclosed in the three images. In addition to the image redundancy, results of a feature point, grid point and edge matching are

combined in a least-squares matching method to check for consistency and for further refinement of the extracted 3D point cloud.

Generation of a DSM by simultaneous processing of more than two overlapping images has some interesting benefits:

(a) First of all, the image redundancy strengthens the reliability of the image orientation or the process to establish the relationship between 3D terrain points and projected 2D image pixels: during bundle adjustment, points in object space are determined by the best fit of three convergent image rays instead of the trivial intersection of two lines.
(b) Secondly, the probability for a successful match, being a unique and correct solution, increases with the incorporation of more images with different viewing angles. The image redundancy leads to a more robust matching as mismatches and ambiguities, in case of multiple candidates, can be more effectively identified. The so-called trinocular vision (Yachida et al., 1986) offers additional constraints and control to check the correctness of proposed matching solutions.
(c) Thirdly, multi-image matching reduces the presence of occlusion substantially due to redundant image information. Photogrammetric theory argues that the projection of a given terrain point must be clearly visible in at least two images in order to provide relevant elevation information.

Some approaches are reported that employ multiple images, but simply by iterative application of standard stereo matching techniques on individual stereopairs and merging of all stereopair results, for example Okutomi and Kanade (1993) and Raggam (2006). Only a few investigations have been published that process and match all images in the block simultaneously. Research on multi-image matching strategies published in Helava (1988), Collins (1996) and Schlüter (1998) can be cited, as well as Baltsavias et al. (2006) and Zhang and Gruen (2006). The latter two discuss specifically multi-image matching of high resolution satellite imagery but none of them deals specifically with an urban environment.

In a first part the image dataset and test site are outlined. The different phases of the surface model generation process and also strategies for reduction of matching and 3D reconstruction problems are addressed in part two. Next, the results are discussed in terms of a visual and quantitative assessment of the generated DSM. To be able to compare the results, the assessment is also performed on a DSM extracted from the ‘standard’ along-track stereopair, which is part of the triplet. The paper concludes with a brief summary and discussion of the findings.

DATA SET AND STUDY AREA

At the launch of the research project in 2007, IKONOS imagery was considered to be the most suitable for conducting the research because of its high radiometric and geometric capacities. The IKONOS platform, launched in 1999, was the first commercial satellite with a high spatial resolution of 1 m and 4 m in panchromatic (0·45-0·90 μm) and multispectral mode respectively. The one meter

resolution supplies a level of detail that makes the identification and 3D reconstruction of individual structures, such as buildings, possible. The earth observation satellite has a polar, sun-synchronous, 681 km high orbit and is equipped with the Kodak Space Remote Sensing Camera, consisting of 13800 pixels yielding a swath width of 11·3 km at Nadir view.

A triplet is constructed out of an along-track IKONOS stereopair acquired in March 2002 and a third image taken in May 2005. The additional image can be considered as a Nadir, near-vertical image yielding less occlusion. Several criteria are respected for selection of a suitable additional image to extract elevation information, for example overlap with stereo couple, cloud-free acquisition, minimal time interval and complementary stereo constellation in terms of azimuth and elevation angle. Despite of the three years time interval, the 2005 image has been considered as the most optimal candidate regarding the other criteria. Satellite image characteristics are summarised in table I.

TABLE I. Characteristics of the three Very High Resolution (VHR) IKONOS images acquired over the study field.

Image ID	*Acquisition date*	*Elevation angle*	*Collection azimuth*	*Sun elevation*	*Sun azimuth*
Forward	2002-03-01	67·59°	1·6°	39·1°	158·3°
Backward	2002-03-01	75·59°	214·1°	39·1°	158·3°
Nadir	2005-05-16	80·93°	23·5°	65·5°	148·4°

The 2002 along-track pair is characterised by a favourable convergence angle of 35·38° or a base-to-height ratio of 0·67. The combination of first the forward, then the backward 2002 image with the 2005 near-Nadir image has a convergence angle of 14·37° and 23·38° respectively. These relatively small stereo intersection angles reflect a rather weak stereo constellation, but on the other hand they give rise to smaller displacements and less occlusion. The proposed tri-stereoscopic approach combines the benefits of both: a large stereo intersection angle for optimal extraction of clearly identifiable points in open spaces and small intersection angle with less occlusion between an image of the along-track stereopair and the additional 2005 image for height extraction in the vicinity of buildings.

A part of the megacity Istanbul, Turkey is chosen as test field (Fig. 1). It is a city characterised by an intense urban growth, with a prominent densification concentrated along the Bosphorus strait. The study field covers the overlapping area between the IKONOS 2002 and the 2005 imagery. It concerns a densely built-up environment with an area of approximately 60 km^2. The area contains Istanbul's urban core, its historic peninsula and extends to the urban fringe in the North. Geomorphologically the terrain is characterised by a rugged topography with elevation ranging from sea level up to 164 m. Most building types within the study area are irregularly constructed and connected multi-family townhouses and apartment buildings. As a result building size and shape may vary markedly. Most roofs within the city core have a flat base but are stuffed with covered terraces, water tanks, elevator shafts, antennas, etc. The narrow streets and the rugged

landscape also taken into account, makes the area quite challenging to model. At the western fringe of the study area, a zone with large rectangular industrial building types occurs while at the northern fringe a region with more residential, free-standing buildings can be found.

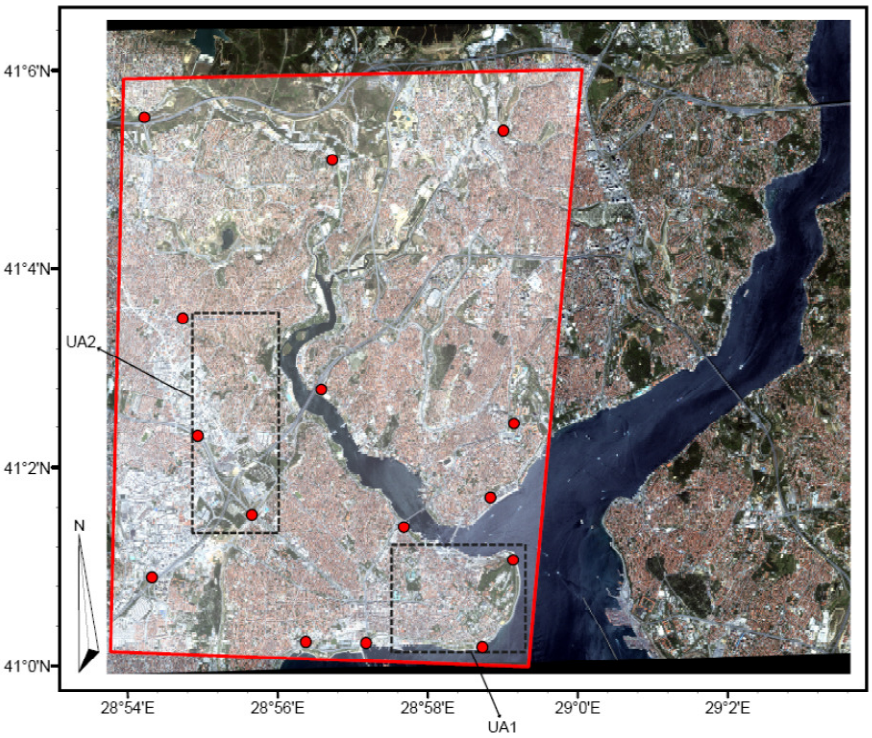

Fig. 1. Overview map with (a) study area, delineated by the red polygon, covering the overlapping area between the three IKONOS images; (b) GCPs, illustrated by red dots, used for image orientation; (c) black dashed polygons indicating 3D building vector validation dataset (IKONOS MS image, 2005-05-16, RGB).

SURFACE MODEL GENERATION

The successive steps of the applied methodology for 3D city surface model generation, based on an IKONOS stereo triplet, are elucidated in following subsections (Fig. 2). The emphasis is especially laid on methods and strategies that cope with the complexity of an urban environment and that reduce the impact of the main distortions, which hamper the 3D reconstruction process.

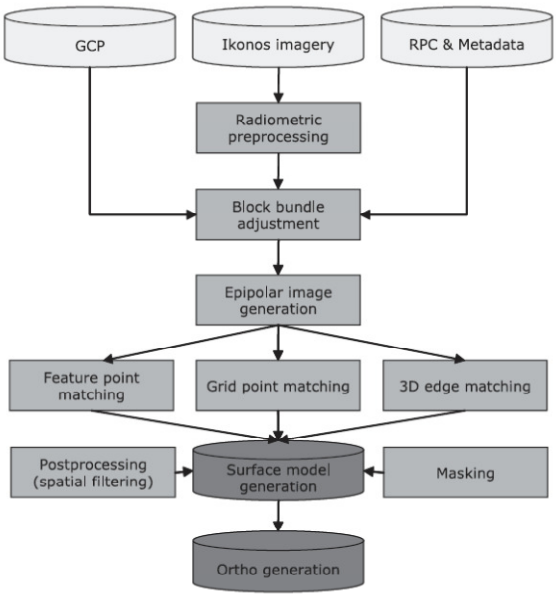

Fig. 2. DSM extraction flowchart.

Radiometric Enhancement and Normalisation

Before processing the VHR imagery, a contrast enhancement and normalisation is executed. Much attention is paid to the image radiometry, particularly to improve texture in shadowed areas, as rich texture is very crucial for the performance of subsequent area-based matching algorithm. Secondly, a normalisation is required as radiometric dissimilarities can occur between multi-temporal images due to different illumination and atmospheric conditions. To enhance the contrast for each image individually and to equalise the radiometric differences, a Wallis filter is applied (Wallis, 1976). As the Wallis filter performs a non-linear, locally adaptive contrast enhancement of an image, it can provide good local contrast at both ends of the dynamic range of grey values. Reduction of surface-independent grey value variation and enhancement of weak texture patterns, such as shadows, leads to a gain in image information and image detail and boosts subsequent image matching process.

Bundle Adjustment for Image Orientation

Rational Polynomial Coefficients (RPCs) are employed to calculate initial values for the image orientation (Grodecki and Dial, 2003). The geopositioning biases caused by the terrain-independency of the RPC sensor model can be further reduced and the model can be further refined through the use of Ground Control Point (GCP) information. Operated with a C-Nav Differential GPS system, 50 well-defined and photo-identifiable GCPs were measured. The system provides sub-10 cm horizontal accuracy and sub-20 cm vertical accuracy. The horizontal

datum is the World Geodetic System 1984 (WGS84) and the vertical datum is mean sea level as determined by the WGS84 Earth Gravitational Model (EGM 96) geoid.

In total, 15 unambiguously identifiable points with known map coordinates were used to establish the relationship between imagery and terrain. Rozycki and Wolniewicz (2007) and Crespi et al. (2009) thoroughly investigated the influence of the number of used GCPs on DSM accuracy, concluding that 10 to 15 well-distributed GCPs need to be used for optimal results. Using a larger number of GCPs did not bring forth significant accuracy improvements, but instead increased the risk of adding an erroneously measured or badly pointed GCP, pulling the overall performance of the model down. The remaining 35 points are used as check points in the accuracy assessment. In order to avoid extrapolation in planimetry and elevation, we aimed to achieve a homogeneous distribution of points. The image coordinates for the GCPs are measured interactively by digitisation on the imagery with a pointing accuracy of one to two pixels. The a priori geometric accuracy for the DSM extraction consists of an overall Root Mean Square Error (RMSE) value of 0·79 m for X residuals, 0·78 m for Y residuals and 2·36 m for Z residuals.

Epipolar Geometry

Prior to surface model extraction, the original images are resampled and rectified to an epipolar geometry based on the orientation parameters of the block bundle adjustment process. An epipolar resampling or normalised image generation applies an affine transformation and aligns the images by a scaling, rotation and translation in Y-direction to yield same geometric properties, enable stereo vision and enhance matching. Y shift between the images is removed, while leaving the parallax or disparity in X-direction unresolved, which can be interpreted as height differences. The epipolar resampling is performed pairwise, with each pair consisting of a selected reference image and one of the search images. Epipolar constraints assist the matching process as they reduce the process of finding homologous points in overlapping images from a two-dimensional to a one-dimensional search algorithm along epipolar lines or the same image row.

Image Matching

One could state that image matching is the core of the photogrammetric process. The matching process implies essentially a description or reconstruction of the 3D environment based on 2D imagery, which is a very difficult and complex problem as it is inverse, ill-posed (Hadamard, 1902; Terzopoulos, 1986) and heavily dependent on the amount of image texture. Image matching algorithms are able to automatically detect homologous points or pixels between all overlapping images, so that corresponding pixels are projections of the same physical point in the object space. Once homologous points are found, a 3D position can be calculated based on the measurement of the disparity or height parallax between the corresponding pixels. This results in a dense 3D point cloud that approximates the surface of the projected objects and from which a DSM can

be interpolated. The matching results strongly determine the level of detail and accuracy of the DSM.

GC³ Multi-Image Matching Strategy. A multi-image matching approach is introduced that fully makes use of the geometric and projective relationships between multiple images and 3D object space. The information enclosed in multiple images, so more than two, and the object space can provide geometric constraints which assist to identify an unambiguous, unique matching solution. The Geometrically Constrained Cross-Correlation or GC³ method is an extension of the standard cross-correlation technique, based on the concept of multi-image matching guided from the object space, so that any number of images can be linked and matched simultaneously (Gruen and Baltsavias, 1988; Zhang and Gruen, 2006).

The applied algorithm works according to a coarse-to-fine hierarchical matching strategy. Image pyramids consist of different versions of an image at exponentially increasing resolutions. The bottom level of the pyramid contains the original image. The matching results of each higher pyramid level are used as approximations in the successive, lower level as guidance to limit the search space. At each level also an intermediate DSM is generated by a constrained Delaunay triangulation (Kallmann et al., 2003), based on the matched primitives, and further refined through the image pyramids.

The GC³ multi-image matching method, applied on the IKONOS triplet, is discussed more in detail below based on the information provided in Fig. 3.

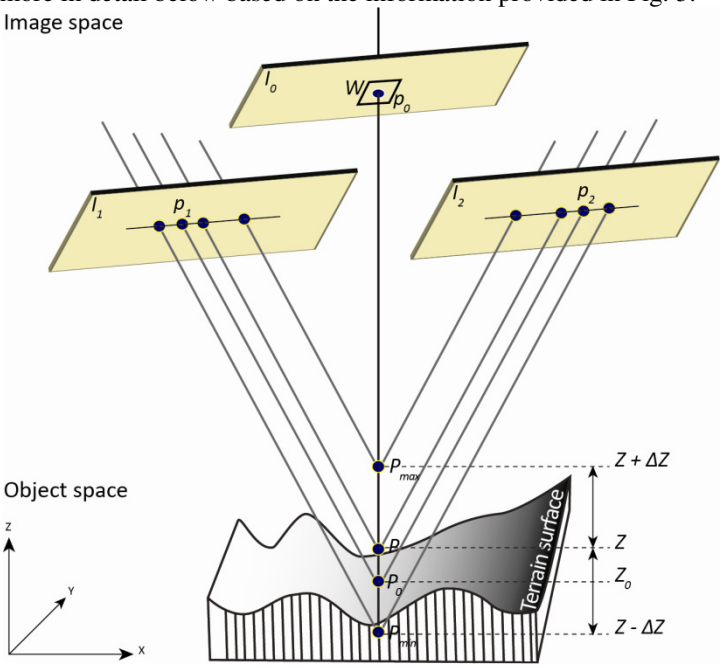

Fig. 3. GC³ multi-image matching strategy, guided from the object space, applied on an IKONOS triplet.

Suppose that both the correct matches or projections p_1 and p_2 on search images I_1 and I_2 need to be found for an image point p_0 on a reference image I_0. Also the correct position P_0 and elevation Z_0 in the object space need to be determined. Based on the calculated image orientation parameters and the RPC sensor model, the image ray through p_0 and the perspective centre of I_0 can be established. The intersection of the image ray with the produced intermediate DSM determines an approximated position P and elevation Z in the object space for the feature point p_0. As elevation is extracted from a rough intermediate DSM, a height error ΔZ is computed based on the height variation in the neighbourhood of the point. The correct position P_0 for the image point p_0 should lie between P_{min} and P_{max}, with height values of $Z - \Delta Z$ and $Z + \Delta Z$ respectively. The GC^3 algorithm implicitly integrates the epipolar geometry constraint for restricting the searching space along corresponding epipolar lines. Perspective back-projection of the elevation interval $[Z - \Delta Z, Z + \Delta Z]$, defined in object space, onto the search images I_1 and I_2 determines the segments of the epipolar lines or search distances on which the correct projections or image points p_1 and p_2 can be found. This terrain height constraint reduces the search space to the image points that lie along a small portion of the epipolar lines and consecutively reduces mismatches.

Then, matching candidates can be determined using a similarity constraint. The Normalised Cross-Correlation (NCC) similarity coefficient is applied to describe the correlation between a template or patch of grey values W around feature point p_0 on the reference image I_0 and each of the correlation-windows around the image pixels on the epipolar line segments in the respective search images I_1 and I_2. The NCC similarity function between the correlation-windows with respect to an absolute height Z_j for p_0 can be defined for each stereopair combination I_0-I_1, I_0-I_2 as:

$$NCC_i(p_0, Z_j) = \frac{\sum_{s \in W} (I_0(s) - \bar{I}_0)(I_i(s_i(Z_j)) - \bar{I}_i)}{\sqrt{\sum_{s \in W} (I_0(s) - \bar{I}_0)^2} \sqrt{\sum_{s \in W} (I_i(s_i(Z_j)) - \bar{I}_i)^2}} \tag{1}$$

with

$$\bar{I}_0 = \frac{1}{m \times n} \sum_{s \in W} I_0(s) \tag{2}$$

$$\bar{I}_i = \frac{1}{m \times n} \sum_{s \in W} I_i(s_i(Z_j)) \tag{3}$$

where i is 1, 2; Z_j is an element of the elevation interval $[Z - \Delta Z, Z + \Delta Z]$; W is the correlation-window around a feature point p_0 in reference image I_0; s is a pixel within correlation-window W; $s_i(Z_j)$ is the corresponding pixel to pixel s in search image I_i; $I_0(s)$ is the grey level value of s in reference I_0; $I_i(s_i(Z_j))$ is the grey level value of $s_i(Z_j)$ in search image I_i; and m and n are the dimensions of correlation-window W.

In contradiction to the standard cross-correlation technique, the GC^3 similarity measure is calculated with respect to an absolute height value Z_j in object space, instead of the relative disparity or parallax between homologous

points in image space. Z_j can be any value between the elevation interval [Z - ΔZ, Z + ΔZ]. In the guidance from the object space lies the true strength of the introduced multi-image matching approach: as the similarity measures are normalised and calculated with respect to an absolute height value Z_j, the NCC functions of the individual stereopair combinations (reference image and one of the search images) can be accumulated so that the ultimate matches p_1 and p_2 get support from all images in the block. SNCC or the mean value of the NCC functions with respect to a certain height Z_j can be defined as:

$$SNCC(p_0, Z_j) = \frac{1}{n}\sum_{i=1}^{n} NCC_i(p_0, Z_j) \tag{4}$$

For each height value Z_j within the interval [Z - ΔZ, Z + ΔZ] an SNCC function can be calculated. Also for each Z_j value, there is a potential match or image point on the search images I_1 and I_2, according to the back-projection onto the respective epipolar line segments. The SNCC function constrains the image matching solution: the correct matches p_1 and p_2 or true height value Z_0 for P_0 will be determined based on the information coming from all images simultaneously, since the ultimate matching candidate will be characterised by a local maximum in the SNCC function (Okutomi and Kanade, 1993). A peak in the SNCC function will correspond to a certain object point P_0 with certain height Z_0 and also to matches or projections p_1 and p_2. These can be defined as the ultimate matching candidates for p_0. The method of using the summed NCC as similarity value, fully benefits of the redundant information and constraints from multiple images and reduces matching ambiguity drastically.

The matching parameters are not user-specified but are determined and fine-tuned progressively based on all data in each pyramid level, following a method proposed by Kanade and Okutomi (1994). This yields an adaptive selection of an appropriate correlation-window size and shape, search distance ([Z - ΔZ, Z + ΔZ]) and threshold for the image similarity value, largely dependent on the sort of imagery and the terrain type. In flat areas or homogenous regions the correlation-window size should increase to detect a statistically matchable pattern and the search distance should decrease while in rough areas or highly textured regions, the patch should become smaller to prevent that the correlation-window extends height discontinuities and the search distance should become longer to deal with relief displacement. Adaptive determination of the matching parameters copes especially with image dissimilarities caused by geometric displacement of objects with certain height.

Multi-Primitive Matching Strategy. In addition to the image redundancy, the matching strategy strives also for redundancy in the matching primitives, being a combination of feature point, grid point and 3D edge matching. This redundancy leads to higher accuracy and reliability of the results since blunders and mismatches can be further detected and removed according a least-squares matching of all extracted primitives (Gruen, 1985). Each of the three employs the GC^3 multi-image matching strategy to detect homologous points (vertices in case of the edge matching).

Feature and grid point extraction generate dense 3D point clouds. Dominant feature points with a kernel of maximum grey level variance are extracted from a reference image based on a local interest operator (Lue, 1988). Instead of the rather randomly dispersed feature points, grid points are uniformly distributed over the whole image. Feature point matching works extremely well in contrast rich areas while grid point matching is especially valuable and complementary in areas with poor texture where conjugate feature points are hard to detect.

3D line segments are complementary to point matching as they comprise more geometric and semantic information than points and are less dependent on image noise. Edges are detected by the Canny edge operator (Canny, 1986) and approximated by connected straight line segments.

3D edge matching is extremely valuable when dealing with an urban case as they assist in modelling surface discontinuities. This is specifically important when buildings and other man-made structures need to be captured and reconstructed. Dense point clouds can describe the terrain surface very detailed but fail to model height discontinuities precisely. The matched edges are taken into account in surface model generation as breaklines or barriers, to avoid smoothing effects where artificial and natural discontinuities in terrain slope appear (for example ridges and building façades). They control the surface behaviour in terms of smoothness and continuity and prevent interpolation operations across discontinuities. In Fig. 4 can be observed that the main shape of structures is estimated quite well by automatically detected edges.

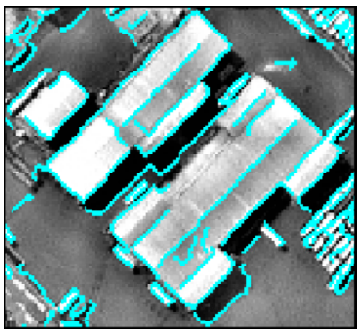

Fig. 4. Binary edge map superimposed on subscene of produced IKONOS ortho-image.

The ultimate matching results on the original resolution imagery are achieved progressively through the image pyramids. 83·3% of 17608174 grid points are successfully matched. In addition, 3919982 feature points and 606221 3D edges are detected and matched.

RESULTS

DSM Interpolation

To generate a connected and continuous terrain surface, the large amount of extracted discrete point measurements need to be interpolated with the extracted

line features serving as breaklines. The generated surface model is processed at a grid size of 3 m. The chosen resolution provides a dense 3D description of the covered surface and leads to the best equilibrium between level of detail and reduction of noise or information overload. In Fig. 5, the interpolated DSM is illustrated for the 60 km^2 wide study area. Also, a zoom in map view on an industrial/commercial site in the southwest and a zoom in perspective view on the historic city centre in the south are provided.

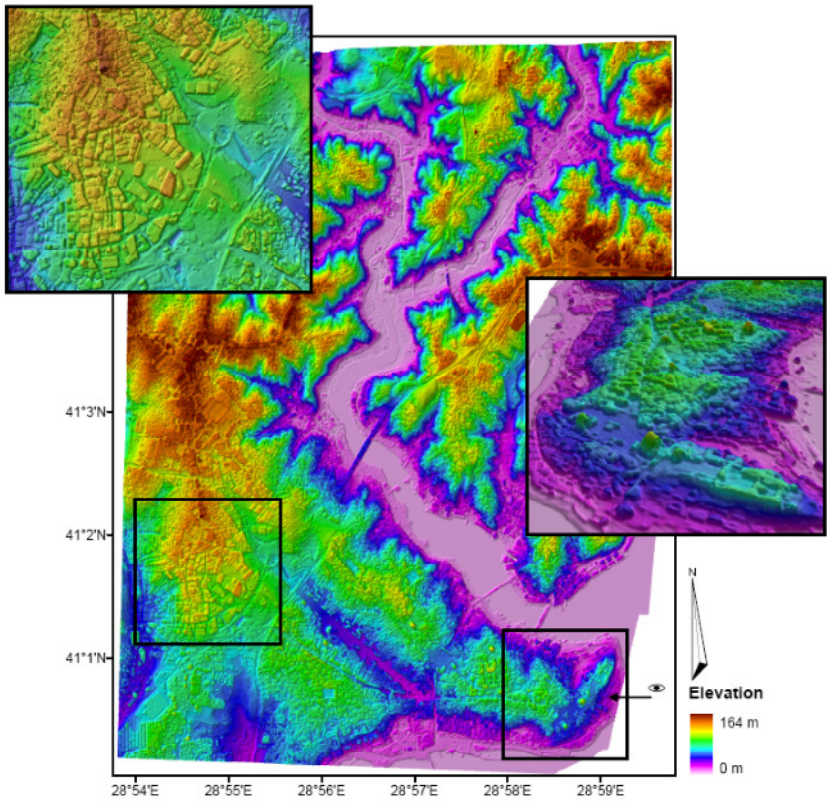

Fig. 5. Extracted 3 m colour-coded DSM, covering the 60 km^2 wide study area. A zoom in map view on an industrial/commercial site in the southwest and a zoom in perspective view on the historic city centre are illustrated additionally.

Qualitative-Visual Analysis

From a qualitative-visual analysis can be observed that the extracted DSM reconstructs the terrain relief, topography and street-house pattern quite veracious and detailed. Large rectilinear and free-standing structures are usually modelled accurately with sharp edges and near-vertical building walls. However, the approach is not robust in the entire dataset. Especially in the very dense urban core, close-set and small buildings are occasionally merged into larger blocks and blob-like shaped building models with fuzzy outlines are still present. The poor modelling of small, close-set buildings is mainly owed to the sparse density of the

3D point cloud as a consequence of residual shadow and occlusion. Also due to the limited resolution of the imagery, too less height data is captured on small dwellings, which hinders an adequate modelling.

Hillshade models are illustrated in Fig. 6 for a subscene of the DSMs, extracted from both the 'standard' along-track IKONOS pair and the IKONOS triplet. The model extracted from only the stereopair shows more flaws and noise, such as spikes and pits, due to a higher percentage of mismatches. Buildings are less well reconstructed and are usually not characterised by a clear outline. Also in many cases (narrow) streets are not detected, resulting in the merge of close-set buildings.

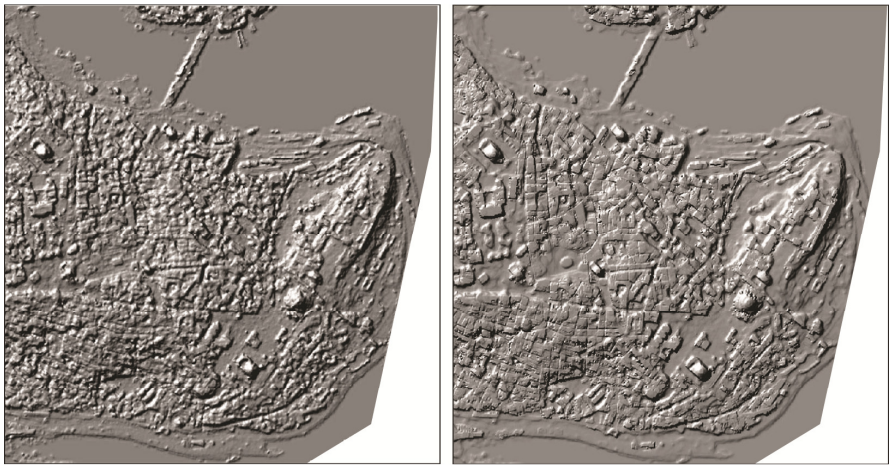

Fig. 6. Hillshade model of the 3 m DSM (subscene of the historic city centre), extracted from the along-track IKONOS stereopair (left) and IKONOS triplet (right).

Quantitative-Statistical Assessment

3D Reference Check Points. A dataset of 35 check points or 'ground truth' is used to check the planimetric and vertical positional accuracy of the extracted models from stereopair and triplet. The check points are independent GPS-measured ground control points, meaning that they are not used in the photogrammetric processing of both models. The generated DSMs are investigated to assess vertical accuracy while produced ortho-images are analysed to check the planimetric accuracy. The residuals for X, Y and Z are expressed in terms of an rmse and summarised in table II for both the 'standard' stereoscopic and tri-stereoscopic approach. In table II, prior to accuracy analysis with independent check points, the geometric accuracy for the image orientation process is given for both models.

TABLE II. A priori geometric accuracy of the image orientation and geometric accuracy analysis based on independent check points.

	A priori geometric accuracy (m)			
	Number of GCP	*RMSEX*	*RMSEY*	*RMSEZ*
Stereoscopic	15	0·68	0·72	2·44
Trinocular	15	0·79	0·78	2·36
	DSM geometric accuracy (m)			
Stereoscopic	35	1·07	1·22	2·61
Trinocular	35	0·81	0·94	2·47

The a priori geometric accuracy predicts the model accuracy well, as the predicted errors are highly correlated with the statistics based on 3D check points. 35 independent check points are used to calculate the rmse for X, Y and Z between measured positions and predicted values of the model. For the trinocular approach, the planimetric positioning error is within one pixel while the vertical error is 2·47 m. The rmse is slightly smaller in X, Y and Z compared to the stereoscopic approach. This can be attributed to the best fit of three convergent image rays instead of the trivial intersection of two lines, for the determination of points in object space.

Although visual comparison of the models shows big improvements for the tri-stereoscopic approach, this does not reflect in the quantitative accuracy check with reference points. The rmse values are slightly smaller for the triplet than for the stereopair, though the differences are not significant. This is due to the fact that the improvements are mainly situated on and in the vicinity of buildings and elevation discontinuities. On the other hand, the assessment represents optimal conditions at check points which are easily identifiable, highly contrasted in the image and located in locally flat areas and open terrain. These requirements are essential to have a sufficient satellite constellation for GPS measurements and to prevent that points are invisible in the imagery due to occlusion. Both models yield comparable results within these non-complex areas.

3D Building Reference. In the second approach, both extracted models are investigated in terms of their ability to model buildings and urban structures. Only vertical positional accuracy of building roof surfaces is assessed. For two parts of the test site a reference database was made available by IMP–Bimtas, consisting of manually measured building heights. Both densely built-up regions are illustrated in Fig. 1 by a black dashed polygon. Zone UA1 covers 5·8 km^2 and is located in the south of the study area. The area contains the city core with mainly blocks of connected multi-family townhouses and historical buildings. Zone UA2, located in the southwest, covers 9·4 km^2 and consists of an industrial area in the south and a new residential area in the northern part.

Buildings with planar roofs were collected: a constant height was collected for each building with Z-map photo software on stereo aerial imagery by digitisation of the outlines of the building at cornice level. The aerial flight has been conducted with an analogue JenOptik LC0030 camera (f= 305 mm) in July

2006. Each image has a photo scale of 1:4500 with a GSD of 9·5 cm and covers almost 1 km^2. The reference dataset is assumed to be of much higher quality than the automatically generated DSMs, because a skilled operator can extract elevation much more accurately than an automated approach, especially on very high resolution aerial imagery.

Firstly, difference maps visualising the 3D error distribution are created by subtracting the rasterised 3D building contours from the produced DSMs according to a pixel-based approach. The histograms of the elevation error maps point out a Gaussian distribution. Secondly, statistical parameters are calculated only for values within the 95% confidence interval, based on these difference maps. The statistics are shown together in table III, describing and summarising the characteristics of the height differences between the generated surface models and the reference dataset. Approximately 11999 buildings are compared in zone UA1 and 41757 buildings in zone UA2. In total, 188138 and 538465 three by three pixels are compared in zones UA1 and UA2, respectively.

TABLE III. Descriptive statistics, calculated from the difference maps "generated surface model minus reference", for the errors within $\mu \pm 2\sigma$.

	Stereoscopic approach		*Trinocular approach*	
	Zone UA1	*Zone UA2*	*Zone UA1*	*Zone UA2*
# Compared pixels	188138	538465	188138	538465
Min dZ (m)	-11·51	-12·89	-6·22	-8·37
Max dZ (m)	5·81	11·16	7·57	7·08
Mean dZ (m)	-3·05	-1·56	0·59	-0·83
St.dev. (m)	3·39	4·56	2·54	3·04
MAE (m)	3·75	3·89	2·06	2·44
RMSE (m)	4·56	4·82	2·6	3·15

The mean height difference or bias error can be interpreted in this context as a signed parameter indicating whether the generated model is biased and if so, in which direction. It is an indicator for a systematic error or more specific for the magnitude of overall under- or overestimation of height in the surface model. As the reference is subtracted from the models, it can be stated that building elevation in the stereoscopic model is underestimated. The bias of the tri-stereoscopic model is much smaller and less than 1 m for both zones. The magnitude of the dZ dispersion or random error of the height differences is measured by the standard deviation. The statistics in table III indicate that the height differences are not spread out over a large range of values but tend to be relatively close to the mean. The standard deviation can also be interpreted as a measure of preciseness. The preciseness of the tri-stereoscopic approach is 0·85 m better for zone UA1 and 1·52 m for zone UA2.

The Mean Absolute Error (MAE) and rmse parameter, both describe the central tendency of the data. The MAE is a linear score, meaning that all the individual differences are weighted equally in the average. The rmse applies a quadratic scoring rule, measuring the average magnitude of the error. As the errors

are squared before they are averaged, the rmse gives a relatively high weight to large errors and makes the measure more sensitive to outliers. For both zones the MAE and rmse values of the trinocular model are significantly better compared to the stereoscopic approach. The errors of the trinocular approach are in the same range as for the assessment with reference points but slightly bigger. This can be explained by the fact that more complex parts of the model, namely building features, are assessed.

CONCLUSIONS

A methodology is presented to extract a surface model and to reconstruct 3D buildings in a highly automated way, solely based on an IKONOS triplet and a set of ground control points. It is tested on a 60 km^2 wide urban environment, consisting of a varying urban morphology and many complex building types. Strategies are proposed at different processing levels to improve DSM generation and to reduce the impact of the main distortions of 3D reconstruction from 2D satellite imagery, being shadow, occlusion, image dissimilarities as a consequence of relief displacement and steep changes in slope. The most prominent strategies are listed below:

(a) As image texture is considered to be the basis for the real strength of an area-based matching algorithm, a radiometric preprocessing is applied to the individual images prior to image matching. The Wallis filter applies a non-linear, locally adaptive contrast enhancement to normalise and enhance image texture, especially in areas with low grey value variance, such as shadowed areas.

(b) The innovative multi-image matching strategy or GC^3 method with implemented SNCC function fully benefits of the geometric and projective relationships between multiple images on the one hand and the object space on the other. Correct matches are not obtained by the analysis of a single NCC function but through the summed NCC values. The SNCC function integrates the similarity constraints, supplied by each stereopair, via guidance from the object space. The image redundancy of a tri-stereoscopic instead of a 'standard' stereoscopic approach strengthens the image orientation, yields a robust matching solution and a resultant higher DSM quality due to additional control and (geometric) constraints (for example the SNCC function constraint), and reduces occlusion substantially. To quantify the reduction in occlusion in case of a tri-stereoscopic approach, a ray-tracing algorithm is applied to the extracted DSM in order to model the occurring occlusion in the different images. Without the redundant image information of the 2005 image, 11·56% of the ground surface is occluded in the 2002 along-track pair by displacement of buildings or other man-made features. If the three images are combined, only 3·89% of the ground surface is covered by occlusion in at least two images and therefore cannot be reconstructed.

(c) In addition to the image redundancy, the approach is also characterised by redundancy in matching primitives. Results of a feature point, grid

point and edge matching are combined by least-squares matching to check for consistency and to refine the extracted 3D point cloud. By comparison of the results of the three matching approaches, remaining outliers and mismatches can be detected and removed more effectively.

(d) Adaptive and self-tuning determination of the matching parameters instead of a fixed set of user-defined parameters copes more effectively with the problem of image dissimilarities/displacement and subsequent mismatches. Adaptive determination takes into account the (local) terrain type and its complexity.

(e) Automatically extracted building edges are added to the 3D point cloud and modelled as breaklines to reconstruct height discontinuities or steep changes in slope, introduced by building façades. Breaklines are taken into account during DSM interpolation to reduce smoothing and rounding of (natural and artificial) discontinuities.

Some conclusions can be drawn from the quantitative assessment of the generated surface models. From DSM accuracy assessment studies of different land cover types, for example Baltsavias (2006), it is well-known that city areas as well as vegetated areas generally yield big elevation errors. According to GeoEye (2006), the IKONOS GEO image product should achieve an overall horizontal positional accuracy of 4 m Circular Error (CE90) and an altimetric accuracy of 5 m Linear Error (LE90). These values correspond with an rmse of 2 m in planimetry and 3 m in altimetry (Dial, 2000). The claimed theoretical boundaries indicate that a feature must be imaged within 4 m of its actual planimetric position with a vertical accuracy of 5 m, at 90% probability or for 90% of the well-defined, photo-identifiable features. For the comparison with reference points, according to table II the attained accuracies for both models are within the theoretical boundaries claimed by the image vendor. The rmse values for X, Y and Z are slightly better for the triplet than for the stereopair, though the differences are not significant. The quantitative-statistical analysis based on 3D building vector data assesses more complex parts of the model, namely building features. For the trinocular approach, significantly better results can be observed. According to table III, the rmse values are in the range of the theoretical values for the trinocular approach. However, rmse values are significantly bigger for the stereoscopic approach.

The benefits and the encouraging results of a tri-stereoscopic approach make it a sound alternative to the more established stereoscopic approaches. The proposed method can be applied on other types of HRSI and on an arbitrary number of images simultaneously without major adaptations. The use of more than three overlapping images would definitely further reduce the impact of remaining occlusion, while the redundant image information would further ameliorate the robustness of the method. Off course one has to consider higher costs when more images need to be purchased and processed. A balance between accuracy and cost-effectiveness needs to be struck in the context of a particular application.

ACKNOWLEDGEMENTS

The Belgian Science Policy Office is gratefully acknowledged for funding the work presented in this paper (SR/00/105). The authors wish to thank the other partners of the MAMUD project (www.mamud.be) and the staff of the remote sensing group at IMP-Bimtas, Istanbul.

REFERENCES

ACKERMANN, F., 1983. High precision digital image correlation. *Proceedings of 39th Photogrammetric Week*, Stuttgart, Germany. 244 pages: 231-243.

ALOBEID, A., JACOBSEN, K. and HEIPKE, C., 2010. Comparison of matching algorithms for DSM generation in urban areas from IKONOS imagery. *Photogrammetric Engineering and Remote Sensing*, 76(9): 1041-1050.

BAILLARD, C. and MAÎTRE, H., 1999. 3-D reconstruction of urban scenes from aerial stereo imagery: a focusing strategy. *Computer Vision and Image Understanding*, 76(3): 244-258.

BALTSAVIAS, E., 1991. *Multiphoto Geometrically Constrained Matching*. PhD Dissertation, Report N° 49. Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland. 221 pages.

BALTSAVIAS, E., ZHANG, L. and EISENBEISS, H., 2006. DSM generation and interior orientation determination of IKONOS images using a testfield in Switzerland. *Photogrammetrie, Fernerkundung, Geoinformation*, 2006(1): 41-54.

BERTHOD, M., GABET, L., GIRAUDON, G. and LOTTI, J. L., 1995. High-resolution stereo for the detection of buildings. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images* (Eds. A. Gruen, P. Kuebler and P. Agouris). Birkhäuser Verlag, Basel. 340 pages: 135-144.

BROWN, M.Z., BURSCHKA, D. and HAGER, G.D., 2003. Advance in computational stereo. *Pattern Analysis and Machine Intelligence*, 25(8): 993-1008.

CANNY, J. F., 1986. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence*, 8(6): 679-698.

COLLINS, R., 1996. A space-sweep approach to true multi-image matching. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, California. 930 pages: 358-363.

CRESPI, M., DE VENDICTIS, L., GRUEN, A., IANNUCCI, G., POLI, D., VOLPE, F. and WANG, X., 2007. DSM extraction from QuickBird and CartoSat stereopairs: quality assessment and comparison. *Proceedings of Conference on Optical 3D Measurement Techniques*, Zurich, Switzerland. 437 pages: 195-203.

CRESPI, M., FRATARCANGELI, F., GIANNONE, F. and PIERALICE, F., 2009. Chapter 4 - overview on models for high resolution satellites imagery orientation. In *Geospatial Technology for Earth Observation data* (Eds. D. Li, J. Shan and J. Gong). Springer, Heidelberg. 572 pages: 63-104.

DHOND, U.R. and AGGARWAL, J.K., 1989. Structure from stereo - a review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6): 1489-1510.

DIAL, G., 2000. IKONOS satellite mapping accuracy. *Proceedings of ASPRS Annual Conference*, Washington D.C.. 8 pages (on CD-ROM).

DPCOR, 2000. *DPCOR user's manual*. Institute of Photogrammetry and Engineering Surveys, University of Hanover, Germany. 12 pages.

FÖRSTNER, W., 1999. 3D-city models: automatic and semi-automatic acquisition methods. In *Photogrammetric Week '99*. (Eds. D. Fritsch and R. Spiller).Wichman Verlag, Heidelberg. 374 pages: 291-303.

GABET, L., GIRAUDON, G. and RENOUARD, L., 1997. Automatic generation of high resolution urban zone digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(1): 33-47.

GEOEYE, 2006. *IKONOS-imagery Products Guide*. Version 1·5. GeoEye, USA. 19 pages.

GRODECKI, J. and DIAL, G., 2003. Block adjustment of high-resolution satellite images described by rational functions. *Photogrammetric Engineering and Remote Sensing*, 69(1): 59-70.

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

GRUEN, A. and BALTSAVIAS, E., 1988. Geometrically constrained multiphoto matching. *Photogrammetric Engineering and Remote Sensing*, 54(5): 633-641.

GUO, T. and YASUOKA, Y., 2002. Snake-based approach for building extraction from high-resolution satellite images and height data in urban areas. *Proceedings of the 23rd Asian Conference on Remote Sensing*, Kathmandu, Nepal. 7 pages (on CD-ROM).

HADAMARD, J., 1902. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 13: 49-52.

HEIPKE, C., 1996. Automation of interior, relative and absolute orientation. *International Archives of Photogrammetry and Remote Sensing,* 31(B3): 297-311.

HELAVA, U., 1988. Object-space least-squares correlation. *Photogrammetric Engineering and Remote Sensing*, 54(5): 711-714.

JACOBSEN, K., 2006. Digital surface models of city areas by very high resolution space imagery. *Proceedings of EARSeL Workshop of the SIG Urban Remote Sensing*, Berlin, Germany. 10 pages (on CD-ROM).

JULIEN, P., 1999. Principles of digital matching. *Proceedings of OEEPE-Workshop on Automation in Digital Photogrammetric Production*, Marne-la-Vallée, France. 475 pages: 161-172.

KALLMANN, M., BIERI, H. and THALMANN, D., 2003. Fully dynamic constrained Delaunay triangulations. Part 4 in *Geometric Modelling for Scientific Visualization* (Eds. G. Brunnett, B. Hamann and H. Mueller). Springer, New York. 488 pages: 241-258.

KANADE, T. and OKUTOMI, M., 1994. A stereo matching algorithm with an adaptive window: theory and experiment. *Pattern Analysis and Machine Intelligence*, 16(9): 920-932.

KRAUS, T. and REINARTZ, P., 2010. Enhancement of dense urban digital surface models from VHR optical satellite stereo data by pre-segmentation and object detection. *International Archives of Photogrammetry and Remote Sensing*, 38(part 1): 6 pages (on CD-ROM).

KRZYSTEK, P., HEUCHEL, T., HIRT, U. and PETRAN, F., 1996. An integral approach to automatic aerial triangulation and automatic DEM generation. *International Archives of Photogrammetry and Remote Sensing,* 31(B3): 405-414.

LAFARGE, F., DESCOMBES, X., ZERUBIA, J. and PIERROT-DESEILLIGNY, M., 2008. Automatic building extraction from DEMs using an object approach and application to the 3D-city modeling. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(3): 365-381.

LEMMENS, M., 1988. A survey on stereo matching techniques. *International Archives of Photogrammetry and Remote Sensing,* 27(B3): 11-23.

LUE, Y., 1988. Interest operator and fast implementation. *International Archives of Photogrammetry and Remote Sensing*, 27(B3): 491-500.

OKUTOMI, M. and KANADE, T., 1993. A multiple-baseline stereo. *Pattern Analysis and Machine Intelligence*, 15(4): 353-363.

RAGGAM, H., 2006. Surface mapping using image triplets: case studies and benefit assessment in comparison to stereo image processing. *Photogrammetric Engineering and Remote Sensing*, 72(5): 551-563.

ROZYCKI, S. and WOLNIEWICZ, W., 2007. Assessment of DSM accuracy obtained by high resolution stereo images. *Proceedings of the ISPRS Hannover Workshop 2007 "High Resolution Earth Imaging for Geospatial Information*", Leibniz, Germany. 6 pages (on CD-ROM).

SCHLÜTER, M., 1998. Multi-image matching in object space on the basis of a general 3D surface model instead of common 2·5D surface models and its application for urban scenes. *International Archives of Photogrammetry and Remote Sensing*, 32(4): 545-552.

TERZOPOULOS, D., 1986. Regularization of inverse visual problems involving discontinuities. *Pattern Analysis and Machine Intelligence*, 8(4): 413-424.

VOSSELMAN, G. and HAALA N., 1992. Erkennung topographischer paßpunkte durch relationale zuordnung. *Zeitschrift für Photogrammetrie und Fernerkundung,* 60(6): 170-176.

WALLIS, R., 1976. An approach to the space variant restoration and enhancement of images. *Proceedings of Symposium on Current Mathematical Problems in Image Science*, Monterey, California. Pages 641-662.

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

WOO, D-M. and PARK, D-C., 2011. Stereoscopic modeling of building rooftop from IKONOS satellite image data. *Proceedings of Conference on Information Science and Applications*, Jeju Island, Korea (South). 5 pages (on CD-ROM).

YACHIDA, M., KITAMURA, Y. and KIMACHI, M., 1986. Trinocular vision: new approach for correspondence problem. *Proceedings of International Conference on Pattern Recognition*, Paris, France. 1300 pages: 1041-1044.

ZHANG, C. and FRASER, C., 2008. Generation of digital surface model from high resolution satellite imagery. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B1): 785-790.

ZHANG, L., 2005. *Automatic Digital Surface Model (DSM) Generation from Linear Array Images*. PhD Dissertation, Report N° 88. Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland. 199 pages.

ZHANG, L. and GRUEN A., 2006. Multi-image matching for DSM generation from IKONOS imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(3): 195-211.