

REFINEMENT OF DIGITAL SURFACE MODELS THROUGH CONSTRAINED CONNECTIVITY PARTITIONING OF OPTICAL IMAGERY

D. Poli^a, P. Soille^a

^a European Commission Joint Research Center, via Fermi 2749, 21027 Ispra (VA), Italy
(daniela.poli, pierre.soille)@jrc.ec.europa.eu

Commission I, WG IV

KEY WORDS: High resolution, very high resolution, Data fusion, Digital surface models, Segmentation

ABSTRACT:

One of the tasks of ISFEREA action at the Joint Research Center (JRC) in Ispra, Italy, is to assess changes in land cover after natural disasters using satellite imagery. To describe the pre-event and post-event scenarios the available heterogeneous data are analysed and eventually fused in order to achieve the most accurate and reliable information. With respect to automatic 3D information extraction, the availability of accurate and detailed Digital Surface Models (DSMs) is a crucial for automatic building detection and subsequent damage estimation. This paper proposes a methodology for the geometric refinement of an arbitrary DSM using high or very high resolution satellite scenes. The main idea is to fuse the DSM and the scene through an advanced hierarchical image partitioning. The method is applied on a dataset over Catalonia, Spain, provided by ISPRS WGI/4 through the project "Benchmarking and quality analysis of DEM generated from high and very high resolution optical stereo satellite data". During this experiment a WorldView-1 quasi-nadir scene was used for the enhancement of the DSM generated by a Cartosat-1 stereopair. The original and final DSM are compared to the Lidar DSM on the same area for quality analysis. After the description of the methodology, the first results are reported and commented.

1. INTRODUCTION

In the last years the number of Earth-observation platforms equipped with very high resolution (VHR) images with ground sample distance (GSD) smaller than 1m has increased. Those sensors are characterized by different geometric, radiometric, spectral and operational specifications, and in most cases allow the acquisition of multi-angle imagery for automatic digital surface models (DSM) generation and 3D feature extraction. One of the tasks of ISFEREA action at JRC in Ispra (Italy) is to assess changes in land cover after events like natural disasters. VHR satellite imagery are generally used for this scope, as they almost cover any area of the planet, in particular remotely located area. To describe the pre-event and post-event scenarios the available data are analysed and eventually fused in order to achieve the most accurate and reliable information. With respect to automatic 3D information extraction, the availability of accurate and detailed DSMs is a crucial issue for automatic building detection and subsequent damage estimation. While the acquisition of VHR stereo scenes can be planned after the event and used to model the post-event 3D scenario, VHR stereo scenes are generally not available at a time of acquisition before the event; on the other hand single VHR scenes are likely available in the archives, as well as DSMs at medium or low resolution from multi-line optical sensors with simultaneous along-track stereo acquisition, like SPOT-5/HRS, Cartosat-1, ALOS-PRISM, or other sources. In this context it is needed to fuse the available data and enhance the surface models using information from single imagery at higher resolution.

This paper proposes a methodology for the geometric refinement of Digital Surface Models (DSMs) obtained from high or medium resolution stereo satellite imagery, using an advanced hierarchical image partitioning of VHR single scenes. The method is described in Section 2 and applied on a dataset provided by ISPRS WGI/4 through the project "Benchmarking and quality analysis of DEM generated from high and very high

resolution optical stereo satellite data" (Reinartz et al., 2010). The data and first results are showed and commented in Section 3.

2. METHODOLOGY

2.1 Related works

The combined use of digital imagery and surface models finds mainly applications in automatic object extraction and 3D reconstruction from aerial data or very high resolution satellite sensors (Arefi, 2009, Baltsavias et al., 1995, Jaynes et al. 1996, Paparoditis et al., 1998, Lu et al., 2002, Tao and Yasuoka, 2002, Li et al., 2008). Few works aim at using satellite imagery for DSM enhancement. In (Maire, 2010) user-defined semantic contents (sea, lakes, building, road, etc.) are extracted with a supervised classification in high resolution satellite imagery; then in correspondence of each class the surface is modelled by plane surfaces with geometric constraints given by the topological properties of each class and neighbour regions. This approach was applied, for example, for the enhancement of SRTM-X using a 2.5m ground resolution SPOT-5 scene. The approach proposed by (Kraus and Reinartz, 2010) improves computer vision methods for DSM generation, based on epipolar imagery and dense stereo matching (Scharstein and Szeliski, 2002, Hirschmüller, 2005). One stereo image is segmented and transferred to the disparity map, then for each segment the original disparity map is filled with suitable interpolation of the disparities. A peculiarity of this method is that it does not optimize a DSM using orthoimagery, but starts one step before on the epipolar images during the DSM generation and avoids possible errors contained in orthorectified images.

The approach presented in this paper exploits the optical imagery through segmentation, and then enhances the existing DSM with suitable interpolation functions.

2.2 Segmentation

Given a logical predicate, the segmentation of an image is defined as the partitioning of the image definition domain into non-overlapping regions called segments, such the logical predicate returns true on every segment by false on the union of any pair of adjacent segments (Horowitz and Pavlidis, 1976). Recently, a segmentation technique producing a hierarchical partitioning of the image definition domain under connectivity constraints expressed in terms of logical predicates was proposed (Soille, 2007, Soille, 2008). Given the input constraints, the resulting partition is uniquely defined, contrary to most segmentation techniques. By increasing the threshold levels associated to each connectivity constraint, one obtains series of fine to coarse partitions with the pixels at the lowest level and the whole image definition domain at the highest level of the hierarchy.

Constrained connectivity relies on a dissimilarity measurement computed between each pair of adjacent pixels. The absolute difference between the intensity values was used for all experiments hereafter (see Gueguen and Soille, 2011 and Soille, 2011 for other dissimilarity measures). Two pixels are said to be alpha-connected if there exists a path linking these two pixels in such a way that the dissimilarity between all pairs of adjacent pixels of the path does not exceed the value of alpha. The relation 'to be alpha-connected' is an equivalence relation (reflexive, symmetric, and transitive relation) and therefore it leads to a unique partition of the image definition domain into regions of maximal extent called alpha-connected components.

The constrained connected component of a given pixel is then defined as the largest alpha-connected component of this pixel that satisfies a series of constraints. Any number of constraints may be considered. Given a local dissimilarity defined as the absolute difference, the most natural constraint is defined in terms of a threshold level on the range of the values of the connected components. This constraint is called the omega-constraint and its associated threshold value is the omega threshold. Given the values of the alpha and omega thresholds, alpha-omega connected component of a pixel is defined as the largest alpha'-connected component such that alpha' is less than or equal to alpha and the range of the alpha'-connected component is less than or equal to omega. The omega constraint prevents linking through transitions, a well-known problem of single-linkage clustering.

A typical choice is to consider the same value for the alpha and omega thresholds. By increasing this value, a fine to coarse hierarchy of connected components is obtained. In practice, it is desirable to obtain segmentation as coarse as possible but retaining all structures of interest in the image given the considered application. In this study, since we are interested in buildings and other man-made structures, the parameters need to be selected in such a way that these objects are simplified as much as possible (i.e., ideally are matched by one connected component) but not merged with other objects.

To further favour linking within homogeneous regions while preventing it through transitions, an image pre-processed by the edge sharpening technique described in (Soille, 2010, Soille, 2011) is applied before segmentation. The pre-processing consists in considering as seeds all local maxima of the input image and propagating their values so as to cover the whole image definition domain. The propagation is controlled by the original image values (the smaller the difference between a seed and a non-seeded adjacent point, the greater the speed).

An example of constrained connectivity partitioning is shown in Figure 1. In this figure, the original image is partitioned with increasing values of alpha and omega parameters. To highlight the hierarchical property of the resulting partitions, random

colours were given to the connected components of the finest partition while the colour of a given connected component of a subsequent level is inherited from the largest connected component of the previous level, contained by this connected component.

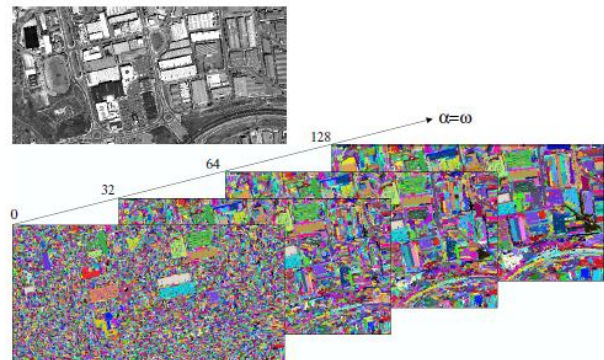


Figure 1. Example of constrained connectivity partitioning of a 512x256 sample of a WorldView-1 image resampled at 2.5m, rescaled on byte data type, and contrast enhanced. The partitions are displayed for the alpha and omega thresholds values equal to 0, 32, 64, and 128 respectively (from fine to coarse partitions)

2.3 DSM enhancement

After the segmentation, the segments are overlaid on the existing DSM and the statistics (mean, sigma, minimum and maximum values) of the heights of the points falling in each segment are computed. The new surface model is calculated from the existing one by imposing that the height values that fall in the same segment are homogeneous with respect to certain mathematic conditions. The new height values of each segment can be the minimum, mean or maximum values of the heights in the segment or the result of planar fitting.

The choice of the mathematic function for the computation of the new DSM is based on the resolution of the input data and the surface characteristics and the DSM application.

3. EXPERIMENTAL RESULTS

In our experiment the algorithms described in Section 2 were used for the enhancement of a DSM generated from a stereopair acquired by Cartosat-1 sensor using a WorldView-1 scene. In the next paragraphs the data and detailed processing will be presented and discussed.

3.1 Data description

The Working Group 4 of Commission I on "Geometric and Radiometric Modelling of Optical Spaceborne Sensors" is providing on its website several stereo data sets from high and very high resolution spaceborne stereo sensors on three areas in Catalonia, Spain, covering urban and rural areas and forests in flat and medium undulated terrain as well as steep mountainous terrain. In addition to these data sets, a Lidar DSM generated by the Institut Cartogràfic de Catalunya (ICC) is provided as reference for comparison (Reinartz et al., 2010). Among the three available regions (Terrassa, La Mora, Vacarisses), the dataset located over Terrassa was chosen for our experiments.

The dataset consists of:

- a subset of a stereopair acquired by WorldView-1 (WV1) on 28th August 2008, composed by a nadir (-1.3°) and off-nadir (33.9°) scene, with ground sample distance of 0.50m and 0.76m respectively; both scenes are panchromatic and have size 10'000x10'000 pixels; Rational Polynomial Coefficients (RPC) are available; for our experiment only the nadir scene is used;
- a stereopair acquired by Cartosat-1 (C1) in February 2008, composed by a backward (-5°) and forward (33.9°) viewing scene, with ground sample distance of 2.5m; both scenes are panchromatic; RPCs are available;
- ground coordinates of points;
- dense 3D point cloud acquired on 27th November 2007 with airborne Lidar.

The area is characterized by urban, rural and forest cover on flat and hilly terrain (Figure 2). The terrain elevation ranges from 200m to 430m.

Considering the ground resolution of WV1 (0.50m) and C1 (2.5m), we defined as objective of our experiment the generation of a DSM at 2.5m GSD. The processing workflow includes the following steps:

- orientation of C1 stereo scenes
- DSM generation at 2.5m grid spacing (DSM_C1)
- orientation of WV1 nadir scene
- orthorectification of WV1 scene using DSM_C1 at ground sample distance of 2.5m
- segmentation of the WV1 orthoimage
- enhancement of DSM_C1 with generation of a new DSM at 2.5m grid spacing.

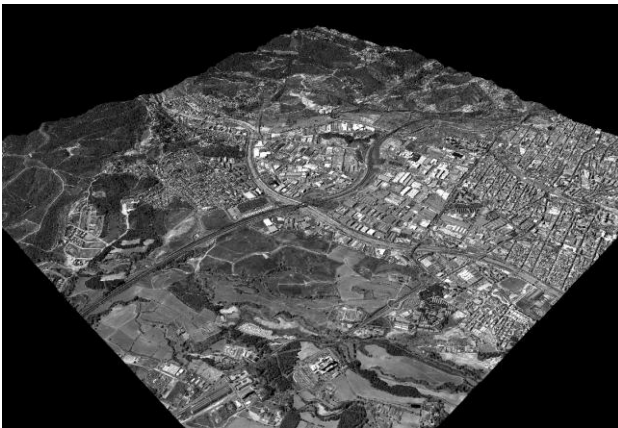


Figure 2. Terrassa area. 3D visualization of textured DSM from WorldView-1 stereo scenes.

3.2 Orthorectification and DSM generation

For the photogrammetric processing of C1 and WV1 scenes the commercial software SAT-PP (SATellite image Precision Processing) by 4DiXplorer AG was used (Zhang, 2005).

The C1 stereopair was oriented by estimating the correction of the available RPC through an affine transformation using nine ground control points (GCPs). Then a DSM was generated using the advanced approach in SAT-PP based on a course-to-fine hierarchical solution with an effective combination of several image matching methods and automatic quality indication (Zhang and Gruen, 2004). To improve the conditions for feature extraction and matching, the images were pre-processed using Wallis filter. Few seed points were measured in the epipolar images in stereo mode to provide the initial approximation of the surface. The DSM was generated with a

regular grid space of 2.5m (DSM_C1, Figure 3); no further editing was applied on the surface model. In the DSM it is possible to distinguish urban areas, but buildings are not well defined from adjacent buildings or roads. In open areas, the surface model follows the terrain shape without outlier.

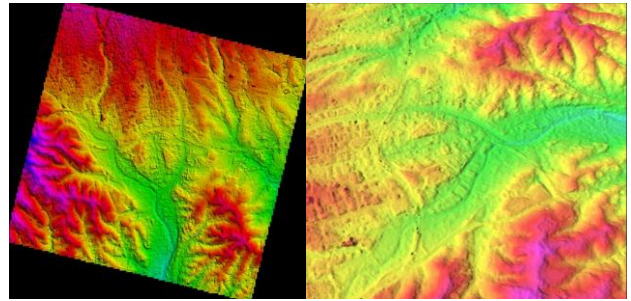


Figure 3. Colour-shaded visualization of DSM generated from Cartosat-1 stereo scenes over Terrassa area. Overview of the full model (left) and zoom on urban and open areas (right).

The nadir WV1 scene was georeferenced, with the estimation of affine correction parameters of the given RPCs by mean of GCPs. Then the orthorectification followed, using the DSM_C1 as elevation model. As a result an orthoimage at 2.5m GSD was produced.

3.3 Segmentation

The WV1 orthoimage was pre-processed with edge-sharpening and segmented. For the segmentation, a value of 50 intensity levels for both the alpha and omega thresholds (see Section 2.2) was found to be optimal, in the sense that it enabled the formation of connected components matching the subparts of the building roofs while avoiding their merging with adjacent objects. An example is given in Figure 4 (a) and (b) on a subset with buildings, main and secondary roads on different levels, a river with bridges, open areas with vegetation and forests.

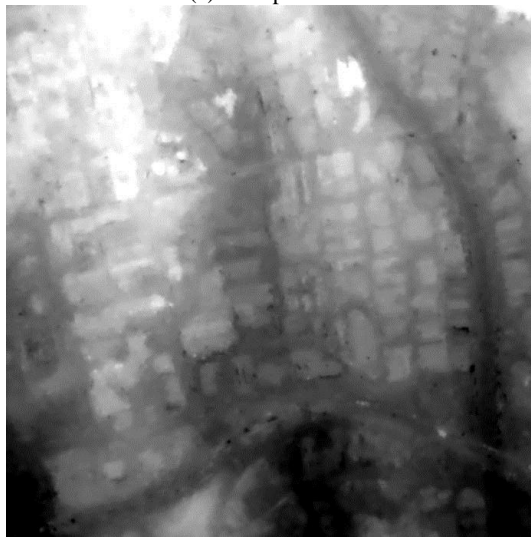
3.4 DSM enhancement and quality analysis

Following the methodology described in (Section 2.3), two DSMs were calculated from the original DSM: for the first one the mean value of the heights in each segment was used (ENH_AVG), while for the second one planar fitting was applied (ENH_PLA). From a visual comparison of the original DSM and the two enhanced DSMs (Figure 4 (b), (e) and (f), respectively) on the subset mentioned above, it is evident that in the enhanced DSMs the building edges are more delineated and objects at different heights, like rivers and bridges, are better defined than in DSM_C1. In homogeneous areas (roof faces, roads, parking areas, fields, and so on) small details due to noise are removed. For the purposes of our research, we focused the analysis on buildings. With respect to the function used to enhance the DSM, there is not a significant difference between average and planar fitting at this resolution, as in DSM_C1 itself it is not possible to recognize shape details, like roof faces, with sufficient accuracy. For the quantitative analysis of the enhanced DSMs, the Lidar DSM was used as reference (Figure 4 (d)). We analysed the differences locally, as at the global level they are influenced by the time differences, and different response on vegetation. Therefore we evaluated the height differences through the surface profile along transects. Three examples are shown in Figure 5.

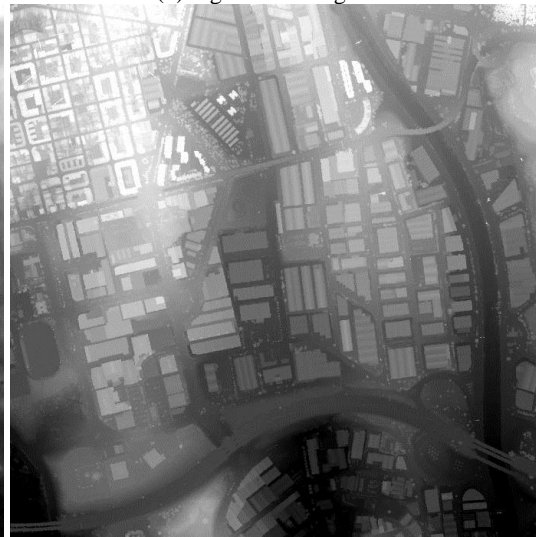


(a) orthophoto

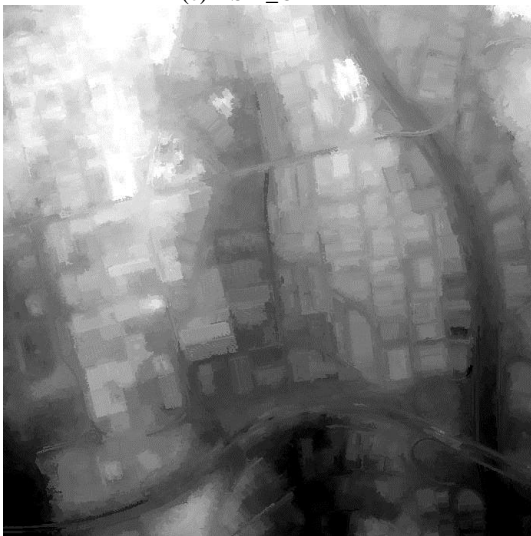
(b) segmented image



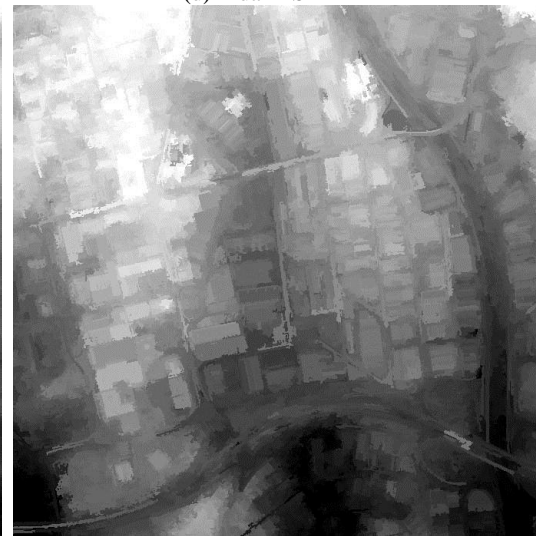
(c) DSM_C1



(d) Lidar DSM



(e) ENH_AVG



(f) ENH_PLA

Figure 4. Segmentation results (left) on a WV1 orthoscene (left), using $\alpha = \omega = 50$ over a subset.

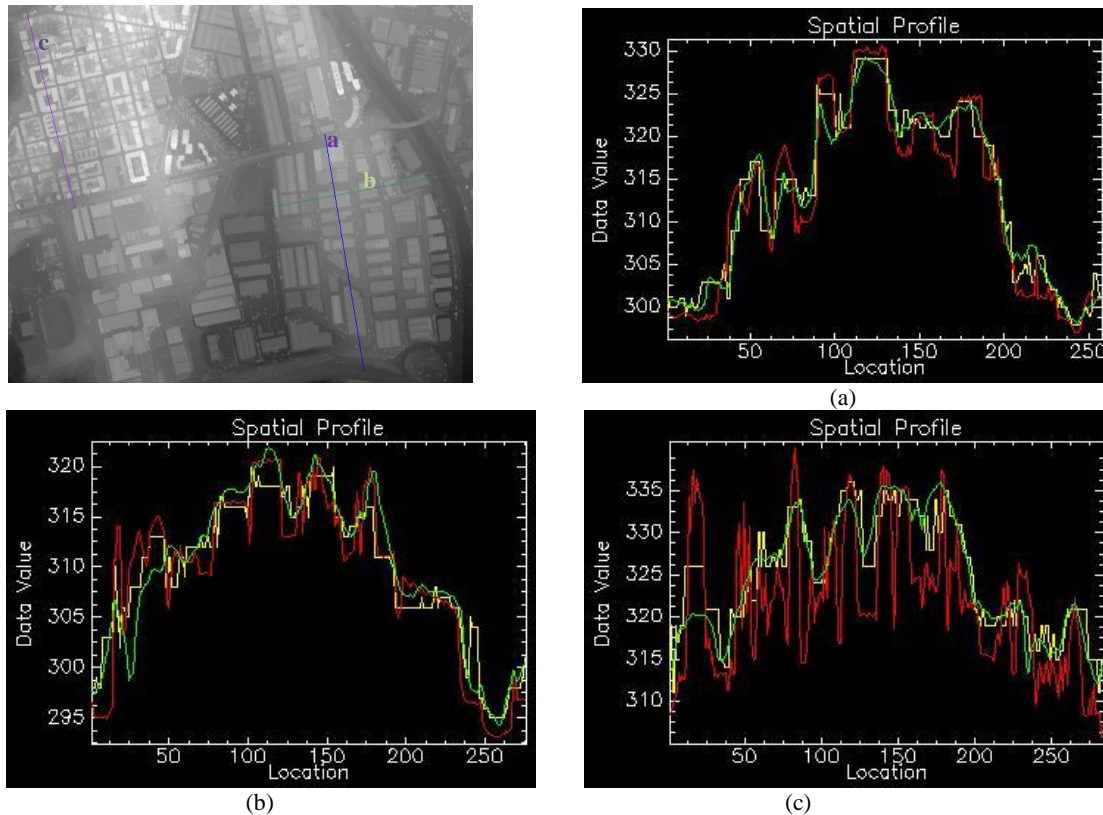


Figure 5. Height profiles of the original DSM (green), Lidar DSM (red) and enhanced DSM with planar fitting (yellow). On the left, the three profile segments are visualized on the Lidar DSM.

The enhanced DSM produces step profiles in correspondence of buildings external edges, which helps the automatic building detection. Obviously the accuracy of the enhanced DSM depends on the accuracy of the initial DSM_C1. So in case of small and closed buildings, like for transect (c), the road height is not modelled in DSM_C1, while it is well defined in the Lidar DSM; the result of the enhancement is a profile that keeps the limitation of DSM_C1 but improves the profile of buildings. The enhancement is not sensitive to the terrain slope.

4. CONCLUSIONS

In this paper we presented an approach for the enhancement of digital surface models using very high resolution satellite images for building extraction; the images are segmented through an advanced hierarchical image partitioning, then the segments are projected on the existing DSM. The new height values are calculated for each segment with planar fitting or constant (average) value of the original DSM heights contained in that segment. The method was applied on a dataset over Catalonia, Spain, provided by ISPRS WGI/4 through the project "Benchmarking and quality analysis of DEM generated from high and very high resolution optical stereo satellite data". During this experiment a DSM was generated by a Cartosat-1 stereopair and a WorldView-1 (WV1) quasi-nadir scene was used for the DSM enhancement. The original and final DSMs were compared to the Lidar DSM on the same area for quality analysis. The first results are promising: the analysis of the profiles on transects over urban areas shows that in the enhanced DSM buildings contours are more sharpen than in the original one, which is favourable for automatic building detection. These first results are promising; further investigations will include the use of multispectral images and the improvement of the computation of the enhanced DSM.

Currently the new height values are calculated as the minimum, maximum or average value of the original heights or by planar fitting on each segment; a combination of these methods based on the surface characteristics will be tested.

ACKNOWLEDGMENTS

Special thanks are given to the data providers for the provision of the stereo data sets, namely: Euromap for the Cartosat-1 data, Digital Globe for the Worldview-1 data and ICC Catalunya for the reference data.

REFERENCES

- Baltsavias, E. P., Mason, S., Stallmann, D., 1995. Use of DTM/DSMs and Orthoimages to Support Building Extraction. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Gruen, A., Kuebler, O., Agouris, P (Eds), Birkhäuser Verlag, Basel, pp. 199-210.
- Gueguen, L., Soille, P., 2011. Frequent and dependent connectivities. *Proceedings of ISMM 2011*, Lecture Notes in Computer Science, Vol. 6671.
- Jaynes, C.O., Collins, R.T., Stolle, F.R., Hanson, A.R., Riseman, E.M., Schultz, H.J., 1996. Three-Dimensional Grouping for Site Modeling from Aerial Images. *Proceedings of Arpa Image Understanding Workshop*, pp. 223-236.
- Hirschmüller, H., 2005. Accurate and efficient stereo processing by semiglobal matching and mutual information. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

- Horowitz, S. and Pavlidis, T., 1976. Picture segmentation by a tree traversal algorithm. *Journal of the ACM* 23(2), pp. 368–388.
- Li, Y., Zhu, L., Shimamura, H., 2008. Integrated method of building extraction from digital surface model and imagery. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing 2008, Vol. XXXVII, Part B3b,.
- Lu, Y. H., Trinder, J., Kubik, K., 2002. Automatic building extraction for 3D terrain reconstruction using interpretation techniques. Techniques, School of Surveying and Spatial Information Systems, University of New South Wales, NSW 2052 Australia.
- Maire, C., 2010. *Image Information Extraction and Modeling for the Enhancement of Digital Elevation Models*. Phd Thesis, Karlsruhe Institute of Technology (KIT), 09.02.2010.
- Paparoditis, N., Cord, M., Jordan, M., Cocquerez, J. P., 1998. Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery. *Computer Vision and Image Understanding*, Volume 72, Issue 2, November 1998, Pages 122-142.
- Krauss, T. and Reinartz, P., 2010. Enhancement of dense urban digital surface models from vhr optical satellite stereo data by pre-segmentation and object detection. In : *The International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences*, Calgary, Canada, Vol. XXXVIII, Part 1.
- Reinartz, P., d'Angelo, P., Krauß, T., Poli, D., Jacobsen, K., Buyuksalih, G., 2010. Benchmarking and quality analysis of DEM generated from high and very high resolution optical stereo satellite data. In: *The International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences*, Calgary, Canada, Vol. XXXVIII, Part 1.
- Scharstein, D. and Szeliski, R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1/2/3):7-42, April-June 2002.
- Soille, P., 2007. On genuine connectivity relations based on logical predicates. In: *Proc. of 14th Int. Conf. on Image Analysis and Processing*, Modena, Italy, IEEE Computer Society Press, pp. 487–492.
- Soille, P., 2008. Constrained connectivity for hierarchical image partitioning and simplification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7), pp. 1132–1145.
- Soille, P., 2010. Constrained connectivity for the processing of very high resolution satellite images. In: *The International Journal of Remote Sensing*, 31(22), pp. 5879–5893.
- Soille, P., 2011. Preventing chaining through transitions while favouring it within homogeneous regions. In: *Proceedings of ISMM 2011*, Lecture Notes in Computer Science, Vol. 6671.
- Tao G. and Yasuoka Y., 2002. Combining high resolution satellite imagery and airborne laser scanning data for generating bare land DEM in urban areas. In: *The International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences*, Volume: XXXIV, Part No.: 5/W3, Kunming, China.
- Zhang, L., 2005. *Automatic Digital Surface Model (DSM) Generation from Linear Array Images*. Ph. D. Dissertation, No. 88, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland.
- Zhang, L., Gruen, A., 2004. Automatic DSM generation from linear array imagery data. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Istanbul, Turkey, Vol. XXXV, Part 3, pp. 128-133.