BENCHMARKING AND QUALITY ANALYSIS OF DEM GENERATED FROM HIGH AND VERY HIGH RESOLUTION OPTICAL STEREO SATELLITE DATA

P. Reinartz[1], P. d'Angelo[1], T. Krauß[1], D. Poli[2], K. Jacobsen[3], G. Buyuksalih[4]

[1]German Aerospace Center (DLR), Remote Sensing Technology Institute, 82234 Wessling, Germany
[2] Joint Research Center (JRC), IPSC, Via E. Fermi 2749, I-21027 Ispra (VA), Italy
[3] Leibniz University Hannover, Nienburger Str. 1 30167 Hannover, Germany
[4] BIMTAS, Evliya Celebi Mah. Mesrutiyet Cad. Eski Tuyap Binasi No:88, Tepebasi-Beyoglu, Istanbul, Turkey
(peter.reinartz, pablo.angelo, thomas.krauss)@dlr.de, daniela.poli@jrc.ec.europa.eu, jacobsen@ipi.uni-hannover.de, gbuyuksalih@yahoo.com

Commission I, WG I/4

KEY WORDS: Optical stereo data, DEM generation, Benchmarking, DEM quality analysis

ABSTRACT:

The Working Group 4 of Commission I on "Geometric and Radiometric Modelling of Optical Spaceborne Sensors" will provide on its website several stereo data sets from high and very high resolution spaceborne stereo sensors. Among these are data from the 2.5 meter class like ALOS-PRISM and Cartosat-1 as well as, in near future, data from the highest resolution sensors (0.5 m class) like GeoEye-1 and Worldview-1. The region selected is an area in Catalonia, Spain, including city areas (Terrassa), rural areas and forests in flat and medium undulated terrain as well as steep mountainous terrain. In addition to these data sets, ground truth data like orthoimages from airborne campaigns and Digital Elevation Models (DEM) produced by laser scanning, all data generated by the Institut Cartogràfic de Catalunya (ICC), are provided as reference for comparison. The goal is to give interested scientists of the ISPRS community the opportunity to test their algorithms on DEM generation, to see how they match with the reference data and to compare their results within the scientific community. A second goal is to develop further methodology for a common DEM quality analysis with qualitative and quantitative measures. Several proposals exist already and the working group is going to publish them on their website. But still there is a need for more standardized methodologies to quantify the quality even in cases where no better reference is available. The data sets, the goal of the benchmarking, the comparison strategy and first very preliminary evaluation results with some data of the selected areas are presented within the paper. The main goal though is to motivate further researchers to join the benchmarking and to discuss pros and cons of the methods as well as to trigger the process of establishing standardized DEM quality figures and procedures.

1. INTRODUCTION

This paper intends to introduce a new benchmarking possibility within the ISPRS community. The main topic of the benchmarking is the automatic generation of Digital Elevation Models (DEM) using optical stereo data from space, which includes, as one of the main processing steps, image matching algorithms. Since processing of optical stereo data is of high interest for many purposes, automatic techniques for image matching and DEM generation have been developed by many institutions to achieve optimum usage of the stereo data sets. A lot of different methods have been developed within the last decades (Reinartz et al. 2006, Passini and Jacobsen 2007) and especially the last years have boosted several new algorithms and methods from computer vision with very interesting results (Hirschmüller 2008, Krauß et al. 2008, d'Angelo et al. 2009).

On the sensor side, several new systems which are able to acquire stereo data from space have been launched in recent years. Especially two kinds of systems are highlighted within this benchmarking exercise. First the along track stereo cameras from Cartosat-1 and ALOS-PRISM: both exhibit a spatial resolution of approximately 2.5 m (GSD) and are specially built for stereo acquisition. These camera systems use two (Cartosat-1) and three (ALOS-PRISM) CCD lines for along track stereo viewing in the same orbit. No special manoeuvres of the satellites are necessary and long stereo stripes can be acquired in a short time span. In contrast to these sensors the new generation of very high resolution (VHR) sensors like Worldview-1 and -2 and GeoEye-1 deliver data with 0.5 m GSD for civil applications. Through their very agile manoeuvring they can also acquire stereo data within the same orbit just using the CCD line combination, by pointing at the same area from two or more orbit positions. This new class of VHR stereo data allow to model also smaller buildings in 2.5D (DSM) or even 3D (object extraction) (Poli et al. 2009). This benchmarking will concentrate first on the 2.5D surface generation but could be used at a later stage also for the benchmarking of automatic object extraction algorithms.

A further goal is to compare methods for DEM quality analysis using qualitative and quantitative measures. Although several methods exist, there is today no standardized methodology for quality analysis and quality figures. The working group will suggest some procedures before and during the benchmarking and is open to receive further input from the ISPRS community.

In order to compare algorithms and methods with the same data sets and with high quality data from airborne campaigns, the described benchmarking data set will be established and can be used by the ISPRS community. Download will be available through the ISPRS web-site via an ftp link. First results will be shown in the workshop "High-Resolution Earth Imaging for Geospatial Information" in Hannover, June 2011.

The paper first introduces the stereo data sets which are open for benchmarking and the reference data set. Some very preliminary results for the DEM generation process are demonstrated and the approach for the benchmarking is shown.

The DEM quality analysis and the comparison strategy are discussed.

2. DATA SETS FOR BENCHMARKING

The new generation of high and very high resolution sensors allow a more detailed generation of Digital Surface Models (DSM). Even for the stereo data with 2.5 m GSD like ALOS-PRISM and Cartosat-1 it is possible to derive a high resolution (5 m spacing) 2.5D model. Especially larger houses are visible in the model, although a real 3D reconstruction of buildings is quite difficult to achieve. Using the data with 0.5 m GSD also smaller buildings and trees can be visualized in the model. The treatment of steep slopes in mountains and water bodies is also a matter of research within the scientific community. Still for all automatic techniques blunders are present and filtering as well as treatment of occluded areas has to be performed in post-processing steps after image matching. Regularization including interpolation and gap filling are further steps in the DSM generation process.

2.1 Test region and area selection

The test region in Catalonia, near Barcelona has been selected due to the availability of several stereo satellite data and a very good reference data set provided by the ICC. In order to be able to investigate different surfaces and for easier data handling and comparison, three smaller areas (each of a 4 km x 4 km size) have been selected according to their properties (see Table 1 and Fig.1).

Test area	Lower left (UTM31N, WGS84)	Area type
1. Terrassa	417400E 4597300N	City, industrial, residential
2. Vacarisses	409100E 4601700N	Wooded hills, quarry, waste dump
3. La Mola	416400E 4608600N	Steep mountainous terrain, forests

Table 1. Position and properties of the selected test areas. The size of each area is 4 km x 4 km.

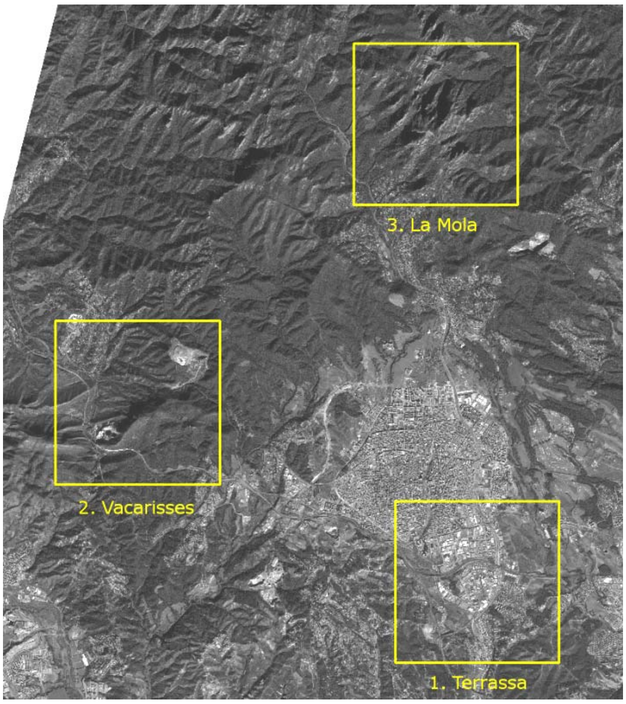


Figure 1. Cartosat-1 image showing the three test areas

2.2 Reference data

The reference data have been provided by the Institut Cartogràfic de Catalunya (ICC). They consist of color orthoimages with a spatial resolution of 50 cm as well as an airborne laser scanning point cloud (first pulse and last pulse) with approx. 0.3 points per square meter. ICC has also provided 143 oriented DMC images with a ground resolution of 25 cm and 60% forward overlap and 50% side overlap, covering all three test sites. We have derived a digital surface model with 50 cm grid spacing (see Fig. 2).

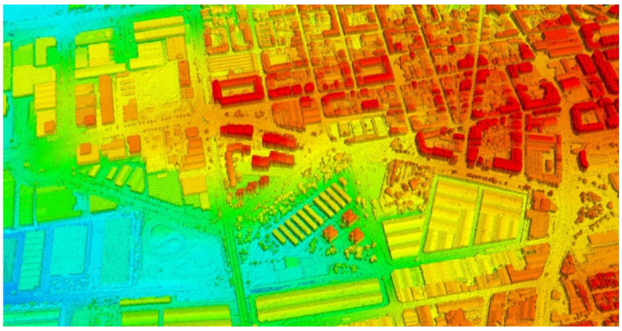


Figure 2. Subset of reference DSM derived from DMC data of Terrassa.

2.3 Datasets

Initially three datasets will be part of the benchmark:

- Cartosat-1, one scene acquired in February 2008.
- ALOS/PRISM, two scenes acquired in April 2008 and July 2009. Unfortunately no cloud free scene for area 3 is available.
- Worldview-1, one L1B stereo scene acquired in August 2008.

In the future, a GeoEye-1 scene and possibly a Worldview-2 scene will also be added.

The datasets provided to the benchmark participants will be cutouts from the original scenes. The scenes are oriented using GCPs derived from the ICC reference dataset. Both the corrected sensor models and the used GCPs will be provided to the participants. RPCs and appropriate corrections will be supplied for Worldview-1 and Cartosat-1. Orbit and attitude data will be supplied for Worldview-1 and ALOS/PRISM. Further investigations are required if RPCs shall be provided for ALOS/PRISM, as some ALOS/PRISM scenes exhibit attitude perturbations which cannot be modelled accurately by the RPCs (Schneider et al. 2008).

The use of the ICC data for both the ground control and the reference for evaluation will minimize offsets between the produced datasets and should allow the estimation of the absolute position error.

3. FIRST PRELIMINARY RESULTS FOR DSM

The GCP collection and orientation of the data is still ongoing, however preliminary results allow production of DSMs from Cartosat-1 and Worldview-1. Due to the preliminary state, only qualitative results are provided in this paper.

DSMs have been computed using the Semiglobal Matching algorithm (Hirschmüller 2008), using an implementation tuned for matching of satellite image pairs. No manual processing has been performed on the resulting DSMs.

Fig. 3 shows a visual comparison between a Cartosat-1 DSM with 5 m grid spacing and a Worldview-1 DSM with 1 m grid spacing. While larger buildings are already contained in the Cartosat-1 DSM, finer details such as the bridges or the residential area, in the lower right corner, can only be extracted from the Worldview-1 stereo scene.

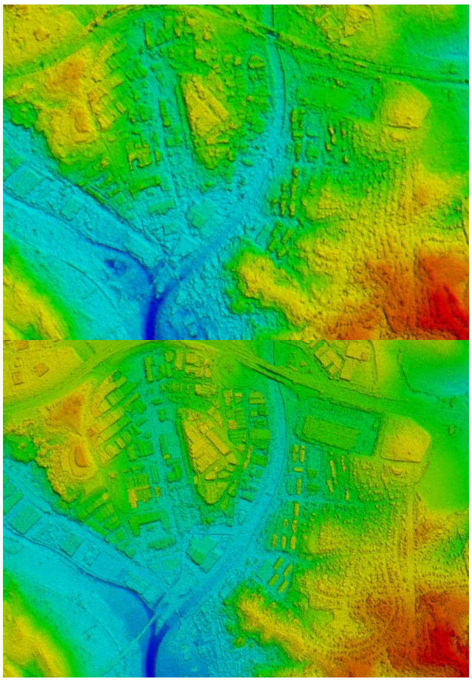


Figure 3. Detail of generated DSM of the Terrassa area. Top: Cartosat-1, Bottom: Worldview-1

A Worldview-1 reconstruction of the complete La Mola area is shown in Fig. 4. The height variation is larger than 600 meters and contains several abrupt cliffs.

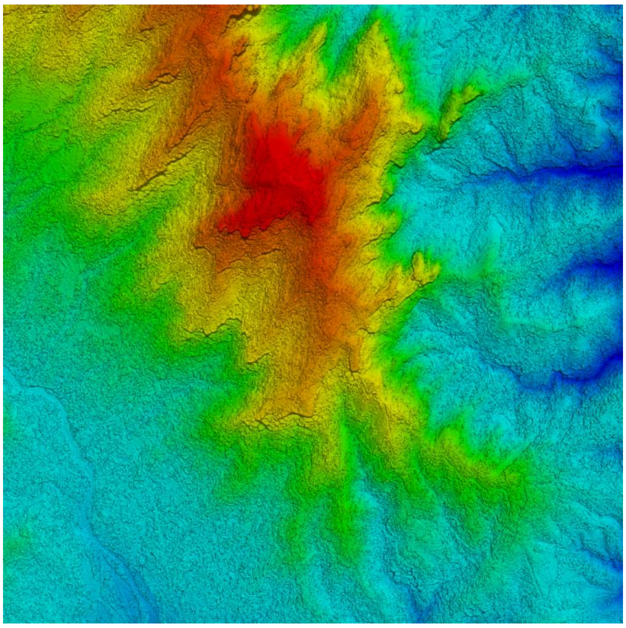


Figure 4. Worldview-1 DSM of the whole La Mola area.

4. DEM ANALYSIS

To analyse a DEM is not a standard operation, it is often missing in papers or done following different strategies. On the other hand the accuracy of a DEM is fundamental to evaluate the performance of the data acquisition sensor and matching algorithm behind, and to assign a quality index to DEM-derived products used in terrain-based applications like orthorectification, mapping, soil-landscape modelling, hydrology, etc. In this section and in the next one we propose qualitative and quantitative measures of a DEM and a standardized methodology for quality assessment using a reference surface model.

4.1 Geomorphologic parameters

Among the common terrain attributes, slope (or inclination), aspect and roughness are the most significant ones for DEM accuracy analysis (Bolstad and Stowe 1994).Slope is defined as the first derivative of the surface and gives the amount of change in elevation in the steepest direction. Aspect indicates the direction that slopes are facing and is defined as the direction of the biggest slope vector on the tangent plane projected on the horizontal plane (Papasaika et al. 2009).

In a general sense, roughness refers to the irregularity of a topographic surface and cannot be completely described by any single measure (Hoffman and Krotko, 1989). In (Papasaika et al. 2009) the measurement of roughness based on entropy is adopted for DEM quality analysis, because as a statistical measure of randomness, entropy can be used to characterize the local variation of the input DEM. An indication of the roughness can be achieved by the derivation of the slope (Jacobsen 2005).

4.2 Vegetation filtering

Digital elevation models from stereo matching as well as from InSAR X- or C-band are describing the visual surface including the influence of vegetation and buildings and are therefore often called surface models or DSM, while reference height models often describe the bare ground. This leads to the problem of DSM-filtering (Passini et al. 2002). The filtering also is required for DSMs with vegetation changing its height. So in advance to the quality analysis a filtering may be required. Of course in dense forest areas the effect of DSM filtering is limited.

4.3 Analysis of buildings in DEM

With the HR and VHR stereo data it is possible to model roughly (2.5D) also single buildings and urban areas (see also Fig. 3). If just two images are present (one stereo pair, sometimes even with large convergence angles > 30 deg.), this modelling or surface generation will surely not lead to exact results due to occlusions and other effects. Also the 2.5 m data will not result in the formation of nice building shapes especially for smaller buildings. Nevertheless it is very interesting to check whether a rough city model can be build with stereo data from space and what accuracy or shortcoming are resulting. Within this benchmark a thorough comparison with the high resolution surface models of the DMC camera (see section 2.2) can be performed. On one side the number of houses found in the DSM can be compared with the houses of the reference DSM, on the other side the shape and height of the buildings can be compared qualitative and quantitatively.

4.4 Blunders

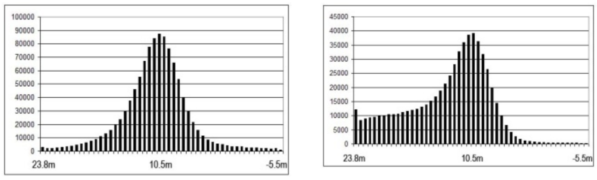
Blunders (large discrepancies outside tolerance) should not be included in the accuracy analysis; they only should be counted by the number or better percentage. Methods for blunder detection and elimination are numerous (e.g. Felicismo 1994). If a coarse DEM is available, this can also be used for some blunder elimination. For the HR and VHR data new methods apply like those shown in this conference by d'Angelo 2010.

5. DEM COMPARISON STRATEGY

In this section a strategy for the analysis of the differences between two DEMs is proposed.

5.1 DEMs alignment

Often height models have not only a bias in vertical direction, also shifts in X and Y may be present due to datum or processing errors. So a constant shift in X, Y and Z direction of the investigated height model to the reference height model by adjustment is optimal. The DEM misalignment can be calculated by matching the two surfaces, like in LS3D software (Gruen and Akca 2007), the Hannover program DEMSHIFT or the DLR software DEM_3D_SHIFT (d'Angelo 2009). If no program for the determination and respecting of the shifts is available, this should be checked by visualisation of the height models and positioning of elements as sharp valleys and similar height discontinuities.
A simple frequency analysis of the height discrepancies of the investigated height model to the reference height model leads to effects of vegetation and to the bias (constant height discrepancies) as shown in Fig. 5.



Frequency distribution just influenced by bias

Frequency distribution of DSM influenced by forest

Figure 5. Frequency distribution of Z-discrepancies

5.2 Error map computation

Once the elevation models are aligned, the height differences provide the vertical discrepancies between the surfaces (2.5D analysis). The Euclidian distances – the shortest distance between an investigated height point and the reference height model – should also be computed to exclude from the analysis the effects of surface-modelling in correspondence of step profiles (Fig. 6).

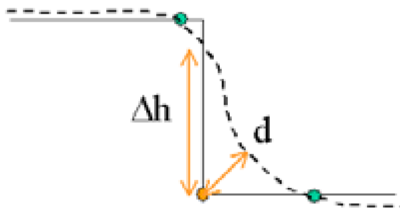


Figure 6. Modelling problems. The true profile is the full blackline, the modelled profile is the dashed line.

For the comparison buildings or generally for urban surfaces a special strategy applies. We suggest performing a comparison on single building level for small, medium and large buildings. The houses found in the DEM and the size as well as the height (e.g. mean or median value for the roof area) can be compared and quantitatively given for certain areas of the city of Terrassa. Also other man-made objects like bridges or streets can be compared on single object or using homogeneity criteria. The error map showing local differences can be used accordingly.

5.3 Accuracy analysis

The root mean square height differences of the investigated height model to the reference one are only a rough figure about the accuracy. At first the root mean square differences without influence of bias are required and in addition the dependency upon the terrain inclination, roughness and aspect should be determined.
As obvious in Fig. 7 usually a linear dependency of the height differences upon the tangent of terrain slope exist, requiring a description of the standard deviation of the Z-component (SZ) by the formula:

$$SZ = a + b*\tan(\alpha) \quad (1)$$

where α is the terrain inclination.

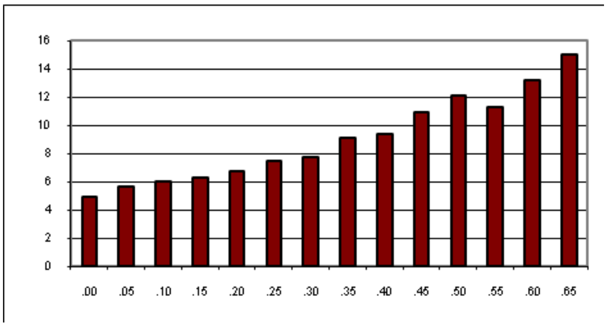


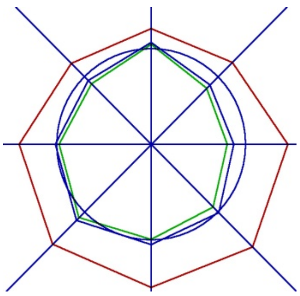
Figure 7. Root mean square height differences SRTM X-band DSM against reference DEM. Horizontal axis: tan(α), vertical axis: RMSE [m]

Some height models have not only a bias but also a scale error in Z, requiring an investigation of bias and scale as

$$Z' = a + b * Z \quad (2)$$

In case of 3D differences, in the formula (1) only the factor b is smaller for the Euclidian distances in relation to the height discrepancies; the influence of the Euclidian distance is significant locally in case of special terrain shapes or surface misalignments.

A linear dependency between height differences and roughness is also expected.
For some methods of height determination, the accuracy may vary with the terrain aspect, in combination with the terrain inclination and flight direction (Fig. 8).



blue circle = RMSZ of all data
blue line = RMSZ of all data, separately for different aspects
green line = RMSZ for different aspects
red line = RMSZ for average inclination

Figure 7. Influence of aspects – example SRTM C-band data

5.4 Land cover

One factor that influences the performance of the different DEM generation technologies is the land cover. For this reason and as mentioned before the whole analysis of the height models should be done separately for the different land cover types, e.g. as open areas, forest areas and build up areas.

6. CONCLUSIONS AND OUTLOOK

This paper intends to show and propose new benchmarking data sets for stereo evaluation of high and very high resolution satellite data, especially for the generation and quality analysis of digital elevation models. The data description shows that with the given GSDs of 2.5 m and 0.5 m even small objects like buildings can be at least partly seen in the generated DEMs. A quality analysis in comparison to the reference DEM (which is as well a surface model) is proposed and it is shown that it will be dependant on the land cover, mainly the classes: urban area, forest and open area. Therefore the quality analysis, which will be qualitative and quantitative, and the comparison strategy will depend on these land cover classes.
On the ISPRS website a link to the stereo data including ancillary data like RPCs and ground control will be given and some description for the benchmarking regarding quality figures and comparison strategies. First results of the benchmarking are envisaged to be shown at the ISPRS Com I Hannover workshop in spring 2011.

ACKNOWLEDGEMENTS

We would like to thank 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, ESA/JAXA for the ALOS-PRISM data and ICC Catalunya for the reference data.

REFERENCES

Bolstad, P.V., Stowe, T., 1994. An evaluation of DEM accuracy: elevation, slope, and aspect. Photogramm. Eng. Remote Sens. 60, 1327–1332.

d'Angelo, P., Schwind, P., Krauß, T., Barner, F., Reinartz, P. (2009). Automated DSM based Georeferencing of CARTOSAT-1 Stereo Scenes. In: International Archives of Photogrammetry and Remote Sensing (39). IPI-Workshop, 2.-5. Jun. 2009, Hannover, Germany.

d'Angelo, P., 2010. Image Matching and outlier removal for large scale DSM generation. Canadian Geomatics Conference and Symposium of Commission I, ISPRS, these proceedings.

Felicísimo, Angel M., 1994. Parametric statistical method for error detection in digital elevation models. ISPRS Journal of Photogrammetry and Remote Sensing, 49(4): 29-33.

Gruen A., Akca. D., 2007. "Least Squares 3D Surface and Curve Matching", ISPRS Journal of Photogrammetry and Remote Sensing, 59(3):151-174.

Hirschmüller, H., 2008. Stereo processing by semi-global matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(2), pp. 328 – 341.

Hoffman, R. and Krotkov, E., 1989. Terrain Roughness Measurement fiom Elevation Maps. SPIE Vol 1195 Mobile Robols IV.

Jacobsen, K., 2005. Analysis of SRTM Elevation Models, EARSeL 3D-Remote Sensing Workshop, Porto, on CD and http://www.ipi.uni-hannover.de.

Krauß, T., Lehner, M., Reinartz, P. (2008). Generation of coarse 3D models of urban areas from high resolution stereo satellite images. In: International Archives of Photogrammetry and Remote Sensing , 37 (B1) , pp. 1091-1098. International Society of Photogrammetry and Remote Sensing. 21. ISPRS Congress on Photogrammetry and Remote Sensing, Beijing, China. ISSN 1682-1750.

Papasaika, H., Poli, D., Baltsavias, M., 2009. Fusion of Digital Elevation Models from Various Data Sources. The International Conference on Advanced Geographic Information Systems and Web Services, February 1-7, Cancun, Mexico.

Passini, R., Betzner, D., Jacobsen, K., 2002. Filtering of Digital Elevation Models, ASPRS annual convention, Washington 2002.

Passini, R., Jacobsen, K., 2007. High Resolution SRTM Height Models, ISPRS Hannover Workshop 2007, IntArchPhRS. Vol XXXVI Part I/W51 + http://www.ipi.uni-hannover.de/

Poli, D., Wolff, K., Gruen, A., 2009. Evaluation of Worldview-1 stereo scenes and related 3D products. ISPRS Hannover Workshop 2007, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVIII-1-4-7/W5.

Reinartz P., Müller R., Lehner M., Schroeder M., 2006, “Accuracy Analysis for DSM and orthoimages Derived from SPOT HRS Stereo Data using direct georeferencing”, ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 60/3, pp. 160-169.

Schneider, M., Lehner, M., Müller, R., Reinartz, P. (2008), Stereo Evaluation of ALOS/PRISM Data on ESA-AO Test Sites - First DLR Results. In: Proceedings of ALOS PI 2008 Symposium . ESA Communication Production Office. ALOS PI Symposium, Rhodos (Greece). ISBN 978-92-9221-228-5. ISSN 1609-042X.