

Journal of
Applied Remote Sensing

**Iterative approach for efficient
digital terrain model
production from CARTOSAT-1
stereo images**

Hossein Arefi
Pablo d'Angelo
Helmut Mayer
Peter Reinartz

Iterative approach for efficient digital terrain model production from CARTOSAT-1 stereo images

Hossein Arefi,^a Pablo d'Angelo,^a Helmut Mayer,^b and Peter Reinartz^a

^a Remote Sensing Technology Institute, DLR—German Aerospace Center,
D-82234 Wessling, Germany

hossein.arefi@dlr.de

^b Bundeswehr University Munich, Institute for Applied Computer Science,
D-85577 Neubiberg, Germany

Abstract. This paper proposes a new algorithm for automatic digital terrain model (DTM) generation from high resolution CARTOSAT-1 satellite images. It consists of two major steps: generation of digital surface models (DSM) from stereo scenes and hierarchical image filtering for DTM generation. Automatic georeferencing, dense stereo matching, and interpolation into a regular grid yields a DSM. In the second step, the DSM pixels are classified into ground and nonground regions using an algorithm motivated from gray-scale image reconstruction to suppress unwanted elevated pixels. Nonground regions, i.e., 3D objects as well as outliers are iteratively separated from the ground regions. The generated DTM is qualitatively and quantitatively evaluated. Height profiles and comparisons between the generated DSM, derived DTM, and ground truth data are presented. The evaluation indicates that almost all nonground objects regardless of their size are eliminated and appropriate results are archived in hilly as well as smooth residential areas. © 2011 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: [10.1117/1.3595265](https://doi.org/10.1117/1.3595265)]

Keywords: digital terrain models; digital surface models; segmentation; semi-global matching; geodesic dilation; image reconstruction.

Paper 10104R received Jul. 14, 2010; revised manuscript received Mar. 24, 2011; accepted for publication May 3, 2011; published online Jun. 1, 2011.

1 Introduction

Digital elevation models (DEM) are fundamental information for many geographic information system applications. The height information used for DEM generation can be extracted using different methods and systems. Stereo image matching is a common technique for the generation of digital surface models (DSMs) in remote sensing and photogrammetry. Stereo pairs of satellite or airborne images acquired from different viewpoints are employed to surface model generation based on digital image matching. Another method is airborne laser scanning or LIDAR, which provides very detailed and accurate height information regarding the ground surface. An interferometric synthetic aperture radar can also be used to extract height information, but the quality in built-up areas is rather poor compared to the other above-mentioned methods.

In May 2005, India launched its IRS-P5 satellite carrying the CARTOSAT-1 instrument, which is a dual-optics two-line along-track stereoscopic pushbroom scanner with a stereo angle of 31 deg and resolution of 2.5 m. The CARTOSAT-1 high-resolution stereo imagery is well suited for the creation of DSMs. The semi-global matching (SGM) algorithm¹ is used to perform dense stereo matching. After outlier removal and transformation into the Universal Transverse Mercator (UTM) projection, a DSM with 5-m grid spacing is obtained.

The DSM produced by automatic stereo matching represents visible surface, including all topographic objects available in the scene; i.e., it is a DSM corresponding to the pixels on top of the visible objects, including buildings and trees as well as some mismatched regions, which cannot be avoided.² To generate a digital terrain model (DTM), which is a digital representation of the ground pixels, all nonground pixels have to be eliminated from the DSM. This includes the effects of 3D topographic objects as well as noise or outliers. Classification of the points into ground and nonground is the basis for DTM generation.

Before describing some of the known filtering techniques, the common problems faced while generating an appropriate DTM will be explained. In most algorithms, the basic objective of processing the height data for DTM generation is to extract the objects with a large discontinuity to their neighboring regions. According to this, the geometric criteria are defined for processing pixels based on a local neighborhood. Filtering algorithms that operate on grid-based DSMs, i.e., height images, require a filled DSM without gaps. In this case, spatial interpolation is employed to rasterize and create the filled DSM. However, interpolation results in a loss in precision particularly for the transition points from a ground to a nonground pixel, e.g., building edges. In this case the transition area will be smoothed with the smoothness depending on the interpolation utilized.

In very hilly or mountainous residential areas where 3D objects such as buildings are located on the hills, separating the nonground from the ground regions is difficult. In such areas large discontinuities can also arise for sharp hills. Therefore, for very hilly residential areas the generation of an appropriate DTM needs additional sources of information.

In addition to the effects of interpolation and slope of the terrain, the filtering algorithm has to deal with outliers or gross errors in the height data. Most filtering techniques that are based on the minimum or maximum values in the local neighborhood are very sensitive to outliers and, therefore, tend to produce unwanted results if outliers are present.

Accordingly, an efficient filtering technique should be independent of the following parameters:

- (a) terrain type: city or forested, hilly or smooth,
- (b) outliers or gross errors,
- (c) interpolation errors,
- (d) size of the 3D objects.

Various techniques and filtering methods have been proposed to generate DTMs from DSM data. A DTM is generated from LIDAR data recorded in wooded areas³ based on a gray-scale morphological opening.⁴ The lowest points within a window of a given size are first detected by the opening. The points in this window falling within a band above the lowest elevation are considered as ground points and weighted surface interpolation using these points is applied to compute the DTM. A conclusion of this work was that the size of the structuring element⁴ (SE) used for opening is a critical parameter for which there is no single optimal value. Therefore, the use of multiple openings with different sizes for the SE was suggested. Kraus and Pfeifer⁵ introduced another algorithm for DTM generation in wooded areas based on linear prediction. They also start with an approximation of the ground surface. The distances from the ground surface to the measured points are used to define weights, which are employed in computing the DTM based on linear prediction. If the height residual within the surface interpolation is above a certain threshold, the corresponding point is classified as nonground and eliminated from the surface interpolation.

A method is described for DTM generation based on progressive densification of a triangular irregular network⁶ (TIN). The idea is to fit a surface from below to the point cloud. In every iteration, the surface is allowed to fluctuate within a certain range, and points from the point cloud are added to the TIN. This is iterated until no further low ground points can be added. The approach has been built into the TerraScan software package.⁷

A slope-based filtering method is proposed for separating nonground points from ground points.⁸ A point is classified as a terrain point if there is no other point in its vicinity to which

the height difference is larger than an allowed maximum difference. The method assumes that the terrain slopes do not exceed a certain threshold and the features having slopes above the thresholds are due to nonground objects. This assumption limits the approach to terrain with gentle slopes. A modified slope-based filtering was proposed based on varying the threshold value according to the terrain slope.⁹

A filtering strategy based on iterative trend elimination and linear prediction to remove nonground regions from the DSM is described in Ref. 10. A program called DTMCOR has been developed at the University of Hannover. The program first locates possible spikes or blunders by the introduction of an acceptable minimum and maximum height. The elimination of the trend is carried out using a moving plane.

A grid-based approach is developed for the generation of DTMs from airborne laser data that begins with a coarse grid spacing of 9 m and defines the raster height by selecting the height within the raster element for which 99% of all points are higher.¹¹ The Laplacian of Gaussian⁴ operator in combination with a weight function is used to detect and remove points that are not considered to be ground points.

A progressive morphological filtering method was developed by Zhang et al.¹² with the focus to exclude nonground measurements from LIDAR datasets. The algorithm applies morphological opening and gradually increases the size of the SE. The resulting elevation differences are used to classify ground and nonground points by a threshold, depending on the SE size. An unsupervised statistical skewness balancing segmentation algorithm is proposed by Bartels and Wie¹³ to detach nonground regions from the ground regions. It is based on the assumption that the naturally measured samples lead to a normal distribution and nonground objects disturb it. Recursive point removal based on skewness is used until a normal distribution is achieved.

A morphological reconstruction algorithm based on geodesic dilation¹⁴ was first employed for segmentation of the high resolution airborne LIDAR data as a basis for separation of ground and nonground pixels.¹⁵ A significant advantage of geodesic filtering is that dilation uses an elementary isotropic kernel, thus, there is no need to specify kernel sizes like in other morphological approaches. The algorithm was tested on a residential area containing objects with different shapes and sizes. The algorithm was enhanced to be used for a hierarchically filtering of the spacious buildings located on hilly terrains using airborne laser scanning data¹⁶ as well as DEM produced from stereo satellite images.¹⁷

In this paper, an iterative algorithm for separating ground and nonground regions is proposed, which is based on a hybrid morphological reconstruction algorithm. The algorithm is independent to the terrain type as well as the size of the 3D objects to be filtered. Within the sequential processing, a classification is carried out to separate nonground and ground regions. This paper is organized as follows. We first describe the algorithm for robust stereo matching used for DSM generation from CARTOSAT-1 images in Sec. 2. In Sec. 3 we give a detailed description of our proposed DTM generation algorithm. Experimental results are presented in Sec. 4. The paper ends with conclusions in Sec. 5.

2 Robust Stereo Matching of CARTOSAT-1 Images

CARTOSAT-1 stereo scenes are provided with a rational polynomial coefficients (RPC) sensor model, as described by Grodecki et al.,¹⁸ derived from orbit and attitude information. The RPC have a much lower absolute accuracy than the ground resolution of approximately 2.5 m. Subpixel accurate ground control points (GCP) were used in previous studies to estimate bias or affine RPC correction parameters required for high quality geolocation. Such highly accurate GCPs are usually derived by a ground survey or from high-resolution orthoimages and digital elevation models. For large scale and continent wide processing, a huge amount of CARTOSAT-1 scenes is required, and the acquisition of the required, highly accurate and GCPs (typically more than five per scene) requires significant resources.

We propose the use of widely available lower resolution satellite data, such as the Landsat-7 Enhanced Thematic Mapper (ETM+) and Shuttle Radar Topography Mission (SRTM) DSM datasets, as reference for RPC correction. The accuracy of these datasets is low compared to the high-resolution CARTOSAT-1 images. The absolute lateral error of ETM+ Geocover is 50 m (LE90). The absolute lateral error of SRTM is between 7.2 m and 12.6 m (LE90, depending on the continent), with an absolute height error of 4.7 to 9.8 m, according to Ref. 19. The traditional method of collecting the lateral position from a reference image and interpolating the corresponding height from the DEM ignores the higher lateral accuracy of the SRTM dataset. The method presented by d'Angelo et al.²⁰ avoids this drawback by using a RPC correction based on DSM alignment, resulting in improved geolocation of both generated DSM and ortho images.

Digital surface models with a resolution of 5 m are derived from dense stereo matching, forward intersection, and subsequent interpolation into a regular grid. First, quasi epipolar images are generated to restrict the search range for the dense stereo matching into one dimension. The SGM algorithm by Hirschmueller¹ is used to perform dense and reliable stereo matching. SGM avoids using matching windows, and is thus able to reconstruct sharp object boundaries that appear washed out when using window-based methods. Instead of strong local constraints in a window, a (semi-) global energy function E is minimized for all disparities (local shifts between the stereo pair) D . SGM performs a semi-global optimization by aggregating costs from 16 directions, which finds an approximate solution to the global energy function E

$$E(D) = \sum_p \left(C(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1] \right). \quad (1)$$

The function C defines the matching cost between the image pixels for each pixel location p in the first image. The simplest cost function is the absolute gray value difference of pixel p and the corresponding pixel in the second image, as defined by the disparity map D_p . The cost function used in this paper is a sum of mutual information¹ and census.²¹ These cost functions adapt to brightness changes in the stereo images and allow matching of images with large viewing angle differences. The second and third terms of E penalize disparity changes in the neighborhood N_p at each position p . T is one if the argument is true, and zero otherwise. The penalty P_1 is added for all disparity changes equal to one pixel. At larger discontinuities (disparity change $>$ one pixel), a fixed cost P_2 is added. This cost function favors similar or slightly changing disparities between neighboring pixels, and thus stabilizes not only the matching in image areas with weak contrast, but also allows large disparity jumps in areas with high contrast.

Minimizing Eq. (1) for two-dimensional neighborhoods, N_p is an NP-complete problem, for which no efficient algorithms exist. In SGM, the minimization is performed by aggregating the cost along eight or 16 paths and tabulating the aggregated costs for all pixels and possible disparities in a 3D array. The effect of the cost aggregation is shown in Fig. 1. The disparity map D is computed by choosing the disparities with the smallest aggregated cost. Matching is performed from first to the second and second to the first image, and only consistent disparities are kept. Finally, small, independent disparity segments are identified and removed as outliers. A detailed explanation of the SGM algorithm is given in Ref. 1.

After forward intersection of the dense stereo matching results and subsequent reprojection into a local coordinate system, the forward intersected points still contain a small amount of blunders due to matching errors in regions with weak contrast or occlusions. To eliminate gross outliers, a check against the SRTM DSM is performed. All points whose height deviates more than 100 m from the SRTM DSM are removed. A second outlier check is done by dividing the data set into small overlapping tiles. The points with minimum and maximum height in each tile are tested using a threshold that is calculated from the height of the points inside the tile by means of Grubbs' test algorithm.²² In each iteration, one point, i.e., either maximum

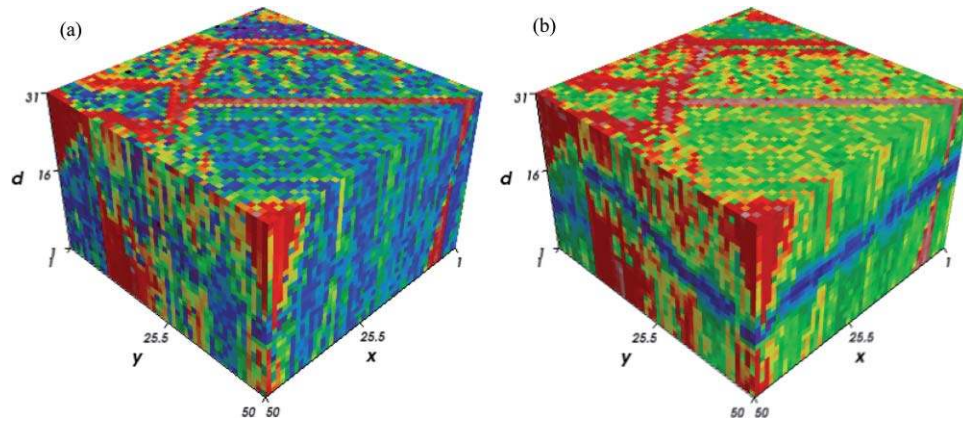


Fig. 1 Raw, pixelwise matching cost C (a) and aggregated costs (b) for all possible matches (row, column, disparity). The pixelwise costs are ambiguous; no clear minimum can be seen on the disparity axis d , whereas, the aggregated costs show a well-defined minimum (blue regions). (Color online only.)

or minimum, is tested. If the point is rejected by the test, the next extrema is validated, until stability is reached, i.e., neither minimum nor maximum is rejected.

The resulting DSM can contain holes in occluded areas and regions where the matching failed or outliers were removed. As most applications require a DSM without holes, a hole filling procedure is required. Small holes in the DSM are filled using inverse distance weighted interpolation, while large holes, often caused by clouds, are filled with SRTM data using the delta surface fill method by Grohman et al.²³

3 Iterative Filtering of Nonground Regions

The DTM generation algorithm of this paper is in first instance motivated by the “fast hybrid gray-scale reconstruction” algorithm described in Ref. 14. The goal of the morphological reconstruction procedure is to suppress unwanted regions with high intensity while preserving the intensity in the regions of interest, which are seeded with “markers” in order to initialize the algorithm.²⁴ Image segmentation based on gray-scale reconstruction is at the core of the algorithm for filtering nonground regions from the elevation image. In this section, a brief description of the concept of gray-scale reconstruction is given as well as a detailed explanation of the proposed implementation of the algorithm by Vincent¹⁴ as “fast hybrid gray-scale reconstruction.” Then, the filtering for eliminating the unwanted nonground regions based on gray-scale reconstruction is described in detail.

3.1 Gray-Scale Reconstruction

“Morphological reconstruction” is a very powerful operator provided by mathematical morphology, which is relatively well known in binary image processing. The operator extracts those connected components of an image, which are marked by another image. The algorithm operates based on a geodesic distance operator.²⁵ It was extended for gray-scale images and has been applied for many image analysis applications such as segmentation and filtering. In photogrammetry, morphological reconstruction has assisted segmentation and filtering of DSM and LIDAR images.^{15,16}

Gray-scale reconstruction based on geodesic dilation employs two input images known as “marker” and “mask” images. The marker image is dilated by an elementary isotropic SE. The resulting image is constrained to remain below the mask image. This means, the mask image

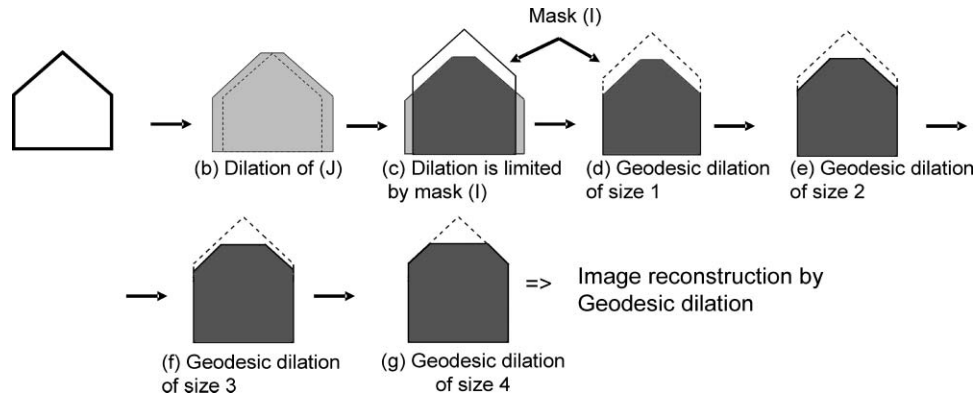


Fig. 2 Geodesic dilation and its comparison to the basic morphological dilation.

acts as a limit for the dilated marker image. This technique is the standard way to perform image reconstruction by geodesic dilation. Basic dilation and point-wise minimum are iterated until stability is reached.

Figure 2 illustrates the difference between geodesic and basic image dilation as well as reconstruction based on geodesic dilation in a profile view of a simple building with a gable roof. The input image in Fig. 2(a), here called marker, is enlarged by dilation, i.e., the gray region in Fig. 2(b), and limited by the mask image (I). The result of geodesic dilation is shown in Fig. 2(d) and a dashed line around it depicts the mask image. If this process, i.e., dilation and limitation by mask, is continued, it stops after four iterations reaching stability. The result provided by this step is called reconstruction of marker (J) by mask (I) using geodesic dilation [cf., Fig. 2(g)].

Accordingly, geodesic dilation and image reconstruction are defined as

$$\delta_I(J) = \delta(J) \wedge I \tag{2}$$

and

$$\delta_I^{(n)}(J) = \underbrace{\delta_I^{(1)}(J) \circ \delta_I^{(1)}(J) \circ \dots \circ \delta_I^{(1)}(J)}_{n \text{ times}}. \tag{3}$$

Equation (3) defines the morphological reconstruction of the marker image (J) based on geodesic dilation $\delta(J)$ [cf., Eq. (2)]. The basic dilation of marker (J) and point wise minimum \wedge between dilated image and mask (I) is iteratively employed until stable. Looking at the reconstructed image of the example depicted in Fig. 2, the upper part of the object, i.e., the difference between marker and mask, is shown as being suppressed during image reconstruction. Therefore, the result of gray-scale reconstruction depends on the height offset between the marker and mask images and accordingly, different height offset suppresses different parts of the object (cf., Fig. 3).

Figure 3 represents the results of geodesic reconstruction of different marker images using the same mask. The first marker is generated by subtracting a small offset $h1$ from the mask image. Image reconstruction suppresses the small top part of the object, here, the building. After subtracting the reconstructed image from the mask image, the upper part of the building, i.e., the ridge, is highlighted. On the other hand, selecting a bigger offset like $h2$ highlights the main part of the building roof after subtracting the reconstructed image from the mask image [cf., Fig. 3(d)]. Thus, different kinds of markers produce different reconstructed images, which may be used in different applications. A small offset can highlight the top pixels of the building that can be used for ridge or local maxima extraction. If the gray-scale reconstruction operates

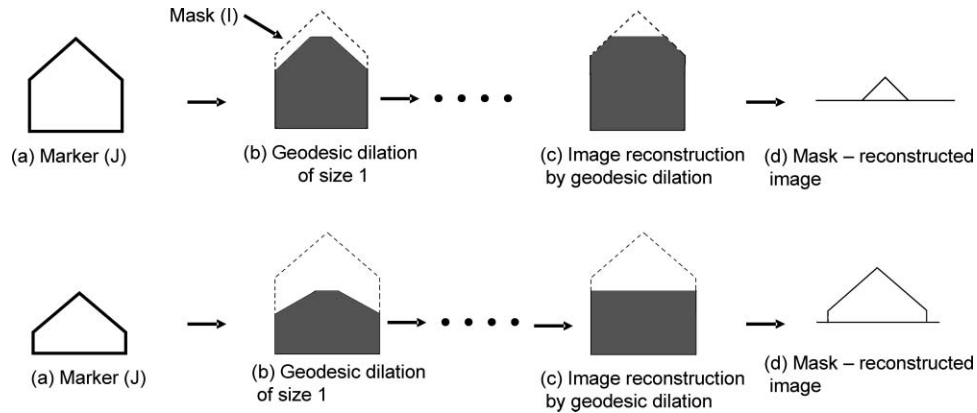


Fig. 3 Geodesic reconstruction of marker images provided by two different offsets of height h_1 and h_2 from the mask image.

ideally, the result following the subtraction of the reconstructed image from the mask image [cf., Fig. 3(d)] is more comparable to what is called the normalized DSM (nDSM) in photogrammetric image processing, i.e., all 3D objects are located on a flat surface. nDSM is used as the basis for segmentation and classification of the nonground objects.

The reconstructed image is formed after iterative geodesic dilation reaches stability. Depending on the size of the image, the number of iterations increases and, therefore, the procedure becomes very time consuming. A fast and efficient algorithm for gray-scale image reconstruction by dilation is proposed, which is termed as “fast hybrid reconstruction algorithm.”¹⁴ The algorithm uses three first-in-first-out (*fifo*) operations:

- *fifo add*(p): Places the pointer to pixel p into the queue.
- *fifo first*(\cdot): Returns the pointer to the pixel at the beginning of the queue and removes the pointer from the queue.
- *fifo empty*(\cdot): Returns “true” if the queue is empty and “false” otherwise.

Accordingly, the fast hybrid reconstruction algorithm operates as follows:

- I : mask image
- J : marker image
- Scan J in raster order:
 - Let p be the current pixel;
 - $J(p) = (\max \{J(q), q \in N_G^+(p) \cup \{p\}\}) \wedge I(p)$
- Scan J in anti-raster order:
 - Let p be the current pixel;
 - $J(p) = (\max \{J(q), q \in N_G^-(p) \cup \{p\}\}) \wedge I(p)$
 - If there exists $q \in N_G^-(p)$ such that $J(q) < J(p)$ and $J(q) < I(q)$ then *fifo_add*(p)
- Propagation: While (*fifo_empty*)
 - $P = \text{fifo_first}()$
 - For every pixel $q \in N_G(P)$:
 - If $J(q) < J(p)$ and $I(q) \neq J(q)$ then $J(q) = \min\{J(p), I(q)\}$ *fifo_add*(q),

where $NG(p)$ is the neighborhood of p on a grid G , with N^+ and N^- denoting the neighboring pixels of pixel p , which are involved for raster and anti-raster scan filtering. \wedge is the point-wise minimum operator¹⁴ (cf., Fig. 4).

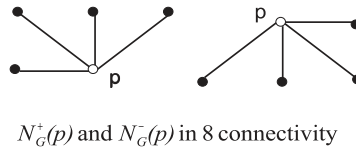


Fig. 4 Neighboring pixels for filtering of point p in raster (left to right) and anti-raster (right to left) order scanning in hybrid geodesic dilation algorithm. In raster order scanning the neighboring points highlighted in $N_G^+(p)$ are points before p and $N_G^-(p)$ are the points after p where, in anti-raster scanning, it is vice versa.

This algorithm implements gray-scale geodesic dilation in two steps; raster and anti-raster direction scanning. It is independent of the size of the objects to be filtered. Therefore, small as well as large objects can be filtered in a similar amount of time.

3.2 Iterative Filtering of Digital Surface Models

The proposed approach for filtering of nonground objects assumes that these objects are higher than their neighboring pixels. Therefore, if a 3D object is not higher than the adjacent pixels on its outside, it will not be extracted using the reconstruction-based filtering.

Image reconstruction by means of geodesic dilation is the basis for segmentation and filtering of nonground regions. One possibility to compare the results of morphological gray-scale reconstruction based on geodesic dilation with the results of a classical opening is to look at the subtraction result of the reconstructed image from the original DSM shown in Fig. 5. The subtraction result of an image and a gray-scale opening of this image is commonly called TopHat filtered image. Applied to surfaces instead of images, the term nDSM is often used instead of TopHat filtered DSM.^{26,27} In the following, we will use the term nDSM.

In the example shown in Fig. 5, the size of the SE of the TopHat filter was 80×80 pixels, which corresponds to 80×80 meters. Defining the size of the SE is one of the critical aspects of basic morphological filtering. Either the SE should be defined by using knowledge about the shape, size, and orientation of the structures, which have to be filtered, or approaches based on a variety of SE sizes have to be developed,²⁸ which eases the dependency on the proper selection of filter kernel sizes. If the TopHat filtered DSM is used for the separation of ground points from nonground points, often times the elevation difference thresholds are applied to the nDSM. These thresholds may be determined by considering the slope, but it remains a crucial aspect to select one or several proper thresholds.

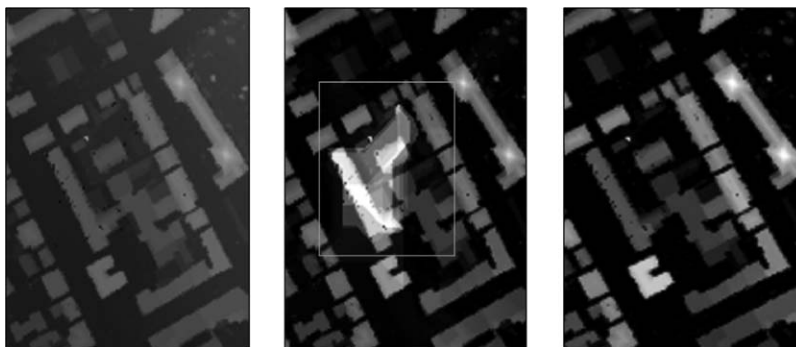


Fig. 5 Comparison of an nDSM produced by TopHat filtering (middle) and gray-scale reconstruction (right).

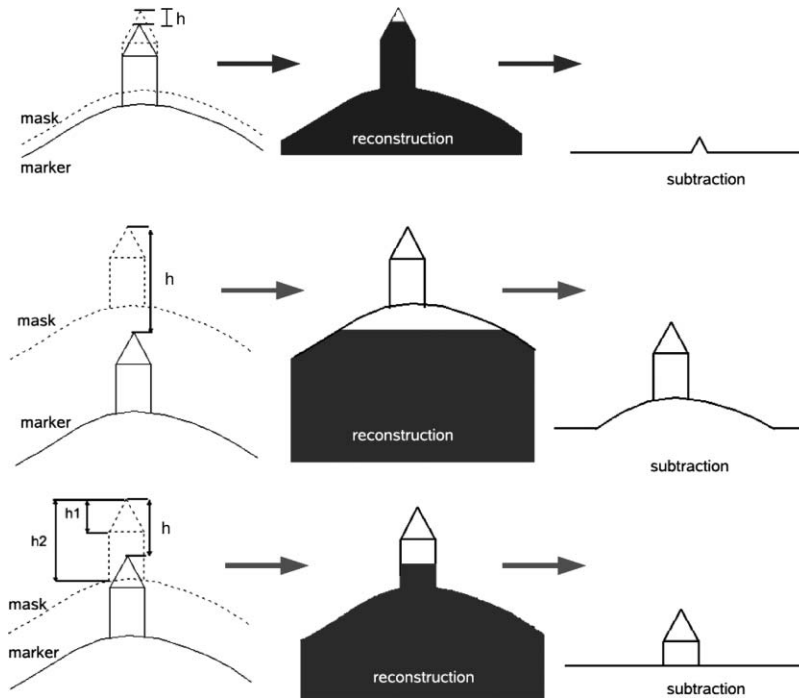


Fig. 6 Selection of an appropriate offset value (cf., bottom) to generate a suitable marker image.

The big advantage of the filtering approach based on morphological reconstruction is that geodesic dilation just employs the elementary isotropic SE. Thus, there is no need to select SE sizes. The successively applied geodesic dilations of size one run until stability is reached. What is more, Fig. 5 indicates that morphological opening tends to produce a surface that lies below the terrain points so that the Tophat filtered DSM is rarely at the zero level [cf., Fig. 5(b) inside white square]. Opposed to this, the nDSM resulting from the reconstruction approach is often at the zero level for the terrain points, thus, separating nonground points from ground points almost naturally. Still, nDSM points with values greater than zero may have to be further classified taking terrain slope or other features into account.

Another advantage of gray-scale reconstruction is that outliers have no negative effect on the filtering. Geodesic reconstruction using marker images produced with any offset value either eliminates or preserves the outliers in the final image; i.e., the process does not enlarge outliers, which is the case for outliers below the ground in opening.

In DTM generation, nonground regions are to be filtered from the original range image. This is similar to the task in gray-scale image reconstruction of suppressing “foreground” pixels. Thus, morphological reconstruction can be utilized in DTM generation as a basic framework. Pixels with high values correspond to 3D objects on the surface, which are to be filtered. The nonground regions contain all 3D elevated artificial objects such as buildings, bridges, vehicles, as well as natural objects such as trees, vegetation, and bushes. This definition does not comprise highly elevated grounds such as hills. Thus, a high value is not sufficient to categorize a pixel as belonging to a nonground region. In order to separate nonground and ground regions using geodesic reconstruction, the algorithm needs to be re-formulated. As shown in Fig. 3, the height of the object to be filtered very much depends on the offset value h for creating the marker image.

Selection of a small offset value results in filtering low objects or just the upper part of high objects, e.g., rooftops, whereas a large offset value might lead to filtering whole objects, e.g., buildings, possibly with parts of the hill below them if they are located on hilly terrain (cf., Fig. 6).

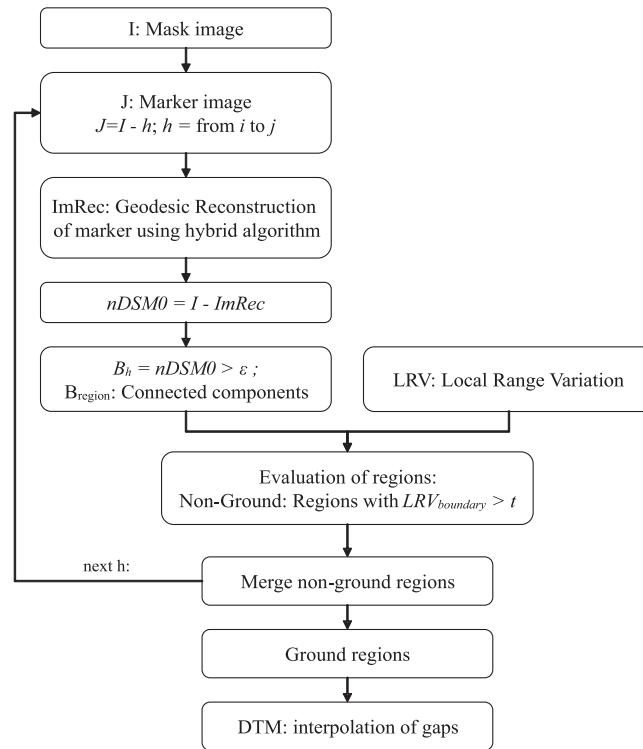


Fig. 7 Workflow of proposed algorithm for DTM generation from DSM produced from robust stereo matching of CARTOSAT-1 images.

A suitable offset for the building displayed in Fig. 3 is a value between h_1 and h_2 (cf., Fig. 6, lower row). Hence, selecting an appropriate offset is critical.

Accordingly, to extract all nonground regions located on any type of terrain, a number of different marker images should be employed. Figure 7 represents the proposed workflow to generate a DTM from the DSM produced from CARTOSAT-1 stereo images. Geodesic reconstruction using different marker images is utilized to iteratively separate nonground regions from the ground regions (cf., Fig. 7).

After producing the reconstructed image using the fast and hybrid algorithm, the result is subtracted from the original DSM (mask image) to generate the initial nDSM (nDSM0). The nDSM in this step is related to the corresponding marker image and it is not the final nDSM. Segmentation by thresholding and connected component analysis creates potential nonground regions, which are evaluated using a feature descriptor. As explained in comparison between geodesic and basic reconstruction techniques, a threshold of zero can be used for segmentation of the nDSM0. To decrease the number of regions to be evaluated, a small value (e.g., 30 cm) is employed for segmentation instead of zero.

To exclude unwanted regions such as hills, the nonground regions are evaluated based on the height discontinuity on the region boundary. The feature that highlights the height discontinuity is called local range variation (LRV). It is generated by subtracting the minimum from the maximum value in 3×3 local windows. The LRV represents the maximum height jump around the pixel. To avoid problems by noise or outliers, the LRV is evaluated using a rank filter. The LRV values of the boundary of each region are sorted, the first and last 5% of the values are eliminated, and the average of the remaining values is assumed to correspond to the height discontinuity of the region.

The LRV feature descriptor highlights the height jumps on the DEM, which is the main criterion for classification of ground, above ground 3D objects, and outliers. As stated, the

3×3 window size is considered to measure the LRV feature descriptor. Accordingly, the quality of LRV decreases with a lower DSM resolution because it cannot represent the height jumps of the 3D objects such as buildings. Therefore, for a very low DSM resolution, the LRV is not able to appropriately differentiate between small hills and spacious buildings due to the same behavior on their pixel boundaries in LRV. Hence, for lower DSM resolution, an additional data source is required for classification of the potential filtered regions.

Only regions having an average height discontinuity along their boundary or more than a predefined value, e.g., 2 m, are classified as nonground. Hills will usually show a very small height discontinuity on their boundaries compared to other 3D objects such as buildings and trees. This procedure is applied for all other offset values h with the number and the values of offsets h computed based on the maximum height discontinuity available in the scene. All nonground pixels are eliminated from the original image so that only ground pixels remain. Finally, the DTM is generated by filling the gaps by means of interpolation.

3.2.1 Marker and mask generation

As shown in Fig. 6, the marker image has a direct effect on the quality of the nDSM0. In this section, we describe an efficient method for generating marker and mask images for each iteration. The offset values h are defined according to the maximum and minimum values of the LRV in the image. The sequential generation of the marker image begins with subtracting the maximum value of the LRV from the mask image as follows:

$$J(i, j) = \begin{cases} I(i, j) & \text{if point } p(i, j) \text{ lies on the border of } I \\ I(i, j) - h & \text{otherwise} \end{cases} \quad (4)$$

$$h = \max(LRV) : -inc : 1,$$

i.e., the border pixels of the marker image are left unchanged compared to the mask and for the rest of the pixels h is subtracted from the mask. To begin, a large h (maximum) has the advantage that regions with smaller h are all included in the regions detected with the bigger h . The iteration continues with producing marker image with smaller value of h , which is computed by subtracting inc from the previous h . Additionally, after each iteration, the pixels belonging to the nonground regions are replaced by a very big negative value, e.g., -9999 . The latter acts as below ground outliers and do not have any effect on the reconstructed image. There are two benefits of this:

1. It makes the process faster, because after extracting the larger parts of 3D objects such as building blocks, the smaller parts will not appear in the next iteration.
2. Particularly in hilly regions with large 3D objects located on them, the replacement means that those large objects are iteratively extracted. As geodesic reconstruction filters objects higher than their surrounding, all pixels on the boundary of an object must have a bigger value than the pixels outside or next to the object. In the DSM, the 3D objects to be filtered, such as buildings and trees, are often higher than their neighborhood, but in suburban hilly regions, it occasionally happens that spacious buildings are situated on steep terrain. An example for such buildings is presented in Fig. 8. In this region, the road next to the building on the left-hand side has a bigger height (327.0 m) than a part of the building on the right-hand side (322.6 m). This means that the building is not completely higher than its neighborhood. In this case, morphologic reconstruction is not able to filter these objects in one step. Therefore, individual parts of the object are iteratively filtered. As shown in Fig. 9, the different parts of the building blocks are filtered in five iterations.

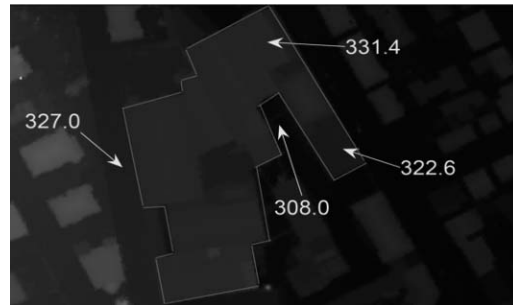


Fig. 8 Large building located on a hill (LIDAR DSM) with heights.

3.2.2 Outlier removal

Two types of outliers namely above- and below-ground outliers are considered in this paper. Above-ground outliers usually consist of a few pixels located above all other 3D objects in their local neighborhood. This type of outlier can be segmented and extracted with the iterative process explained above. They behave like very high objects, such as trees, and therefore, can be eliminated by filtering.

Below-ground outliers are located very much below the ground. To eliminate this kind of outliers, the above-mentioned procedure is repeated using the inverse of the DSM instead of the mask image. Thus, below-ground outliers are converted to above-ground outliers and can therefore be extracted accordingly. Additionally, the criterion for evaluation has to be changed.

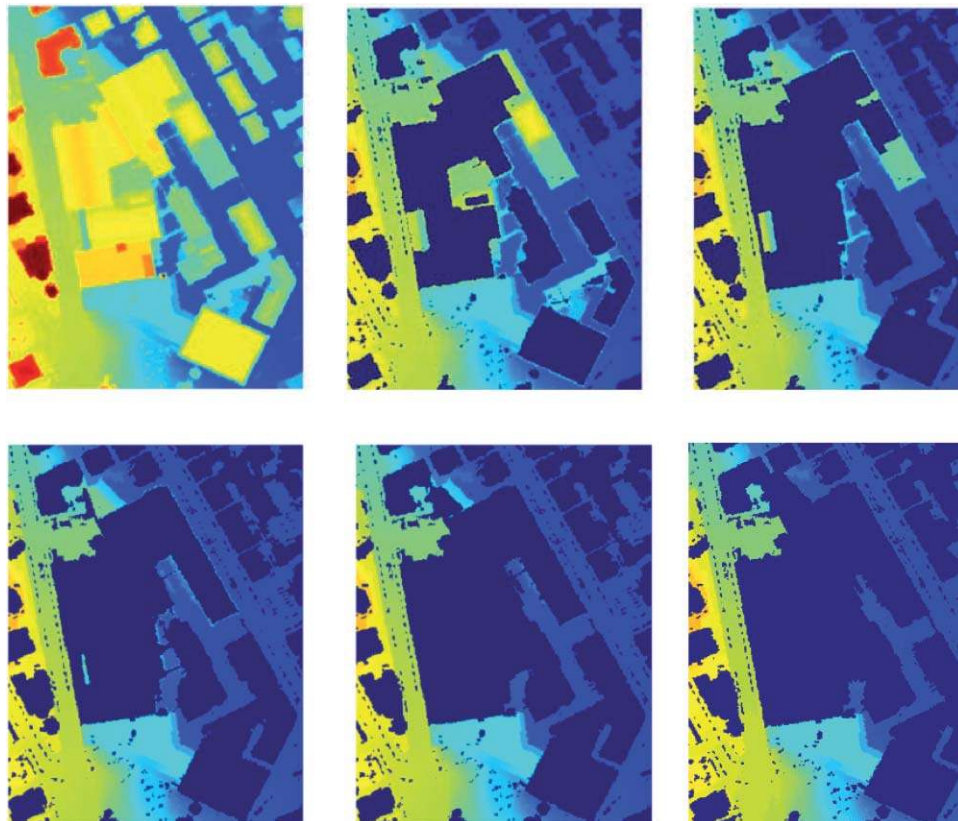


Fig. 9 Iterative filtering of a spacious building located on a hill. Regions classified as nonground are highlighted in dark blue.



Fig. 10 CARTOSAT-1 aft (left) and fore (right) images of the test area.

Since objects, such as entrances to the underground or downward stairs, can also be located below the ground, a high value like 20 m is used to exclude the above objects from filtering.

4 Experimental Results

To evaluate the proposed method for the results of robust stereo matching as well as DTM generation, a hilly residential area also covering parts of the city of Barcelona, Spain, is used as an example. Figure 10 represents the CARTOSAT-1 aft and fore images for the test area.

A DSM is generated using semi-global stereo matching with 5m grid resolution. It contains many gaps produced from mismatched regions. The gaps are filled with SRTM data using the delta surface fill algorithm.²³ Figure 11 shows the created DSM before and after gap filling. The highlighted regions with very bright spots on the left of Fig. 11 represent the mismatched areas. They mostly correspond to bodies of water and flat cultivated fields.

The hierarchical DTM generation algorithm has been employed on the DSM filled by SRTM data. The test data contains large height variations in an area of about 4000 m × 4200 m. The minimum and maximum heights are 21 and 249 m, respectively. The marker images were automatically generated according to the maximum and minimum height variation as defined by LRV (cf., Fig. 12). According to the maximum LRV of 70 m the following offset values were used for marker generation:

$$h = 70 \text{ m} : -inc : 1; inc = 70 \text{ m}/10 = 7 \text{ m}.$$

The number of marker images is always taken as 10 and, accordingly, the *inc* parameter is determined. Then, above-ground objects are iteratively extracted and evaluated by LRV.

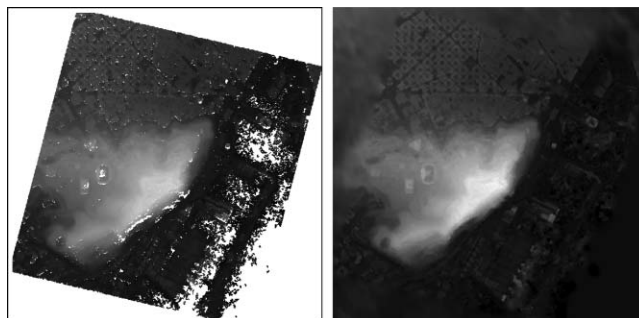


Fig. 11 Left: DSM generated by semi-global stereo matching; right: DSM after filling the gaps by SRTM; height range: [21 m – 249 m], pixel size: 5 m.

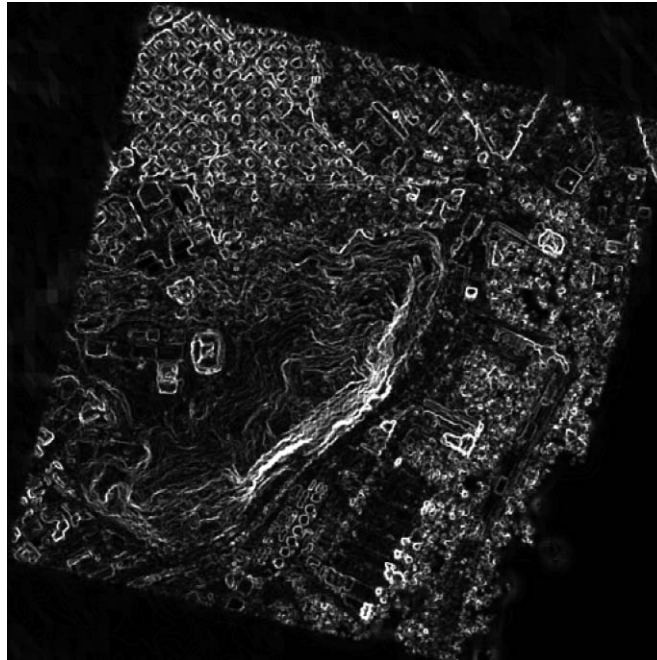


Fig. 12 Local range variation (LRV).

In a similar way, below-ground regions or outliers are extracted and eliminated. After eliminating the detected nonground regions, i.e., above- and below-ground objects, the gaps are filled by means of an interpolation based on the Delaunay triangulation to produce the DTM.

Figure 13 (left) represents the DTM created based on the algorithm proposed in this paper. To evaluate the quality of the produced DSM and DTM, ground truth is provided consisting of a high quality DTM produced from aerial images generated by the Institut Cartografic de Catalunya (ICC) in Barcelona (cf., Fig. 12, right). For the ground truth DTM, a large part of the image corresponding to the harbor and the bodies of water is not available.

Visual comparison between the produced DTM and the reference DTM shows that the proposed algorithm could appropriately eliminate most nonground objects from the DSM, particularly, also objects located on hilly terrain. Looking at the upper part of Fig. 13 shows that buildings located in the flat residential area are not entirely eliminated. This problem is in first instance due to the characteristics of the input CARTOSAT images. The mentioned

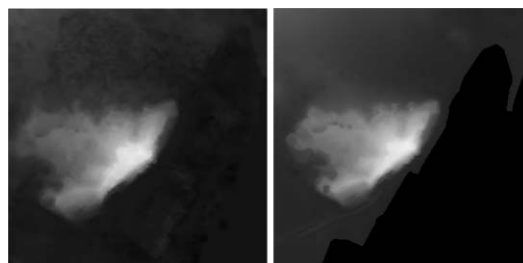


Fig. 13 Generated DTM after filling the gaps by interpolation (left) and corresponding ground truth from ICC (right).

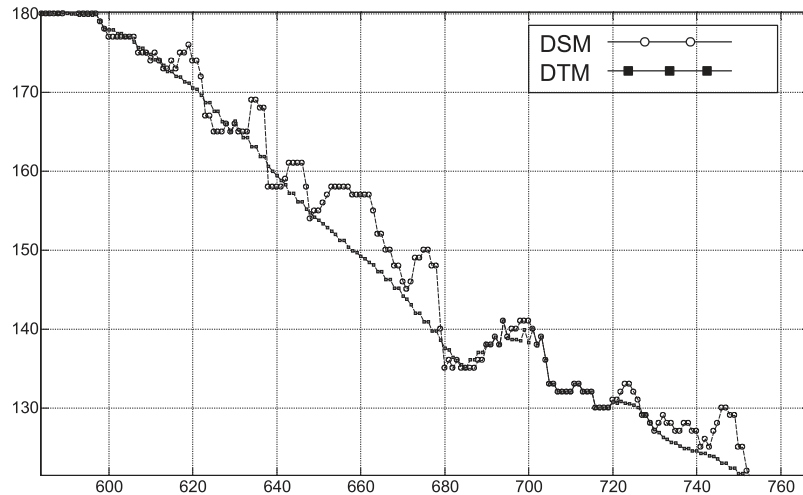


Fig. 14 Profile representing the points of the DSM (dashed line) and of the DTM (solid line).

area consists of many small buildings that are located very close to each other as well as narrow and long roads. Due to occlusions and the resolution of 2.5 m, the CARTOSAT images, in particular the forward-looking one, consist of very few pixels showing the ground in the area between the buildings. Moreover, big shadows between the buildings and on the roads, due to an unfavorable sun angle, additionally decrease the quality of stereo matching in these areas.

In addition to the visual interpretation of the results, DSM, DTM, and ICC height values are given for a profile in Figs. 14 and 15.

Produced DTM pixels after interpolation are plotted as black lines with square signs and compared with DSM values (cf., Fig. 14) as well as with ICC reference pixels (cf., Fig. 15), which are displayed with lines plus circles.

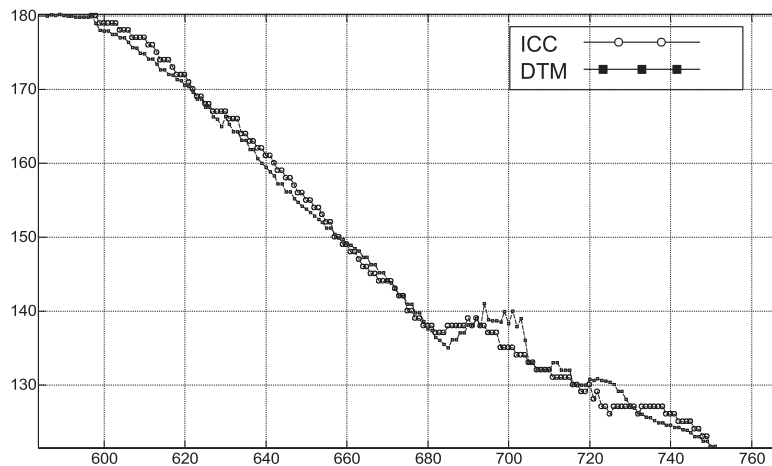


Fig. 15 Profile representing the points of the DTM (solid line) and of the ICC ground truth (dashed line).

Table 1 Statistical evaluation of produced DSM and DTM compared to the ICC reference DTM.

	DTM – ICC	DSM – ICC
Maximum	61.16 m	19.98 m
Minimum	– 34.17 m	– 19.94 m
Mean	5.07 m	1.24 m
Median	2.20 m	0.02 m
Standard deviation	8.20 m	5.66 m
RMSE	9.65 m	5.74 m

For a further assessment, the differences between DSM, DTM, and ICC reference DTM are analyzed. Two differences are produced as follows:

$$\begin{aligned} diff\ 1 &= DSM - ICC \\ diff\ 2 &= DTM - ICC. \end{aligned} \tag{5}$$

From the *diff* images several statistical parameters have been calculated (cf., Table 1). It shows that the mean of the *diff* values is reduced from 5.07 to 1.24 m and the maximum and minimum values are also significantly reduced.

Figure 16 gives two histograms corresponding to the two difference images. The histogram shown in light gray represents the difference between DSM and ICC reference DTM, i.e., *diff 1*, while the other in dark gray corresponds to the difference between DTM and ICC reference DTM.

The histograms as well as the statistical parameters show that the quality of the generated DTM is quite high, particularly because, as discussed above, most of the large differences between DTM and reference DTM correspond to the residential area where the buildings are located very close to each other and, therefore, no ground pixels were generated during stereo matching.

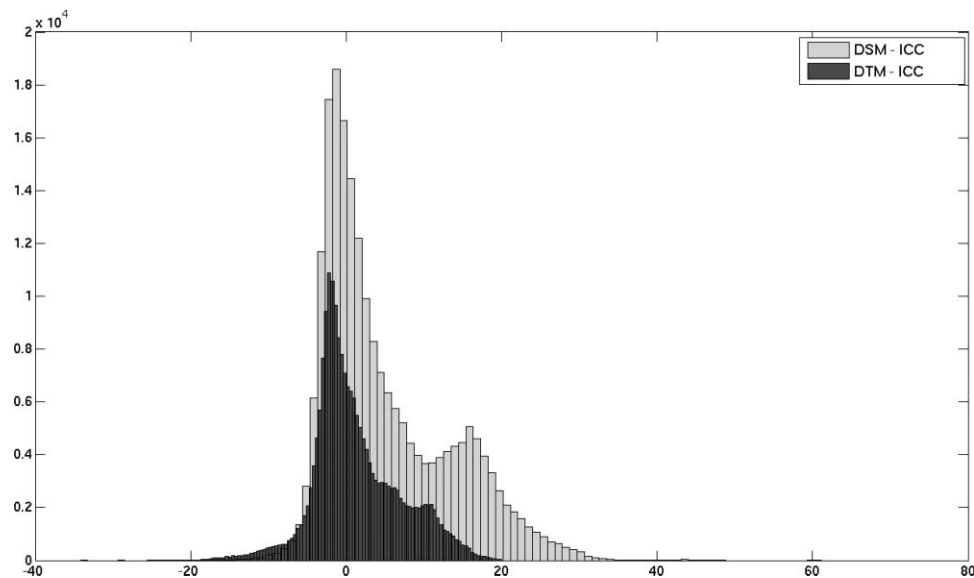


Fig. 16 Histograms representing the difference between DSM and ICC reference DTM (light) and between DTM and ICC (dark).

5 Conclusions

In this paper, a new approach is proposed for automated DTM generation from CARTOSAT-1 stereo images. The SGM algorithm is utilized to perform dense stereo matching. The DSM is generated after outlier removal and transformed into the UTM projection.

Using iterative filtering, 3D above-ground objects as well as below-ground outliers are eliminated from the DSM. The filtering uses the fast and hybrid gray-scale reconstruction algorithm. The algorithm eliminates almost all obvious nonground regions, such as buildings and vegetation areas independent of the shape and size of the objects. Large and small, elongated and compact, as well as high and low buildings are properly eliminated (cf., Fig. 13, left).

Ground objects that have similar characteristics as buildings, such as elevated roads, are also eliminated. The only relevant parameter is the threshold for differentiation between nonground and ground regions. In residential hilly areas, which have been smoothed before DTM generation, height discontinuities for 3D objects are also blurred. In such cases, selecting an appropriate threshold value is critical.

Some regions with very dense large houses, also producing large shadows, are not properly filtered. An improvement for these specific cases and the application to other scenes has to be investigated in more detail in the future.

References

1. H. Hirschmueller, "Stereo processing by semi-global matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(2), 328–341 (2008).
2. K. Jacobsen and R. Passini, "Filtering of digital elevation models," presented at *GIS 2001 Conf. Proc.*, Vancouver (on CDROM) (2001).
3. J. Kilian, N. Haala, and M. Englich, "Capture and evaluation of airborne laser scanner data," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **31**(B3), 383–388 (1996).
4. R. C. Gonzalez and R. E. Woods, "Digital Imaging Processing," Addison-Wesley, Reading, MA (1992).
5. K. Kraus and N. Pfeifer, "Determination of terrain models in wooded areas with airborne laser scanner data," *ISPRS J. Photogramm. Remote Sens.* **53**, 193–203 (1998).
6. P. Axelsson, "Processing of laser scanner data – algorithms and applications," *ISPRS J. Photogramm. Remote Sens.* **54**, 138–147 (1995).
7. Terrasolid, "Processing laser and images, Terrascan software for processing of laser data," <http://www.terrasolid.fi/> [accessed on June 2007].
8. G. Vosselman, "Slope based filtering of laser altimetry data," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **33**(B3), 935–942 (2000).
9. G. Sithole, "Filtering of laser altimetry data using a slope adaptive filter," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **34**(3-W4), 203–210 (2001).
10. R. Passini, D. Betzner, and K. Jacobsen, "Filtering of digital elevation models," in *Proc. of the ASPRS Annual Convention*, Washington DC (on CDROM) (2002).
11. R. Wack and A. Wimmer, "Digital terrain models from airborne laser scanner data – a grid based approach," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **34**(3B), 293–296 (2002).
12. K. Zhang, S. Chen, D. Whitman, M. Shyu, J. Yan, and C. Zhang, "A progressive morphological filter for removing non-ground measurements from airborne lidar data," *IEEE Trans. Geosci. Remote Sens.* **41**(4), 872–882 (2003).
13. M. Bartels and H. Wie, "Segmentation of LIDAR data using measures of distribution," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **36**(7), 426–431 (2006).
14. L. Vincent, "Morphological grayscale reconstruction in image analysis: applications and efficient algorithms," *IEEE Trans. Image Process.* **2**, 176–201 (1993).
15. H. Arefi and M. Hahn, "A morphological reconstruction algorithm for separating off-terrain points from terrain points in laser scanning data," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **36**(3/W19), 120–125 (2005).

16. H. Arefi, J. Engels, M. Hahn, and H. Mayer, "Automatic DTM generation from laser-scanning data in residential hilly area," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **36**(4/W45), 6 pp (2007).
17. H. Arefi, P. d'Angelo, H. Mayer, and P. Reinartz, "Automatic generation of digital terrain models from CARTOSAT-1 stereo images," *ISPRS Hannover Workshop: High-Resolution Earth Imaging for Geospatial Information*, **38**(1-4-7/W5), 6 pp (2009).
18. J. Grodecki, G. Dial, and J. Lutes, "Mathematical model for 3D feature extraction from multiple satellite images described by RPCs," presented at *ASPRS Annual Conf. Proc.*, Denver, Colorado (2004).
19. E. Rodriguez, C. S. Morris, J. E. Belz, E. C. Chapin, J. M. Martin, W. Daffer, and S. Hensley, "An assessment of the SRTM topographic products," Technical Report JPL D-31639 (2005).
20. P. d'Angelo, P. Schwind, T. Krauss, B. Frithjof, and P. Reinartz, "Automated DSM based georeferencing of CARTOSAT-1 stereo scenes," presented at *ISPRS Hannover Workshop: High-Resolution Earth Imaging for Geospatial Information*, **38**(1-4-7/W5), 6 pp (2009).
21. R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. of the Third European Conf. on Computer Vision*, **2**, Springer-Verlag, London, UK, 151-158 (1994).
22. F. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics* **11**(1), 1-21 (1969).
23. G. Grohman, G. Kroenung, and J. Strebeck, "Filling SRTM voids: the delta surface fill method," *Photogramm. Eng. Remote Sens.* **72**(3), 213-216 (2006).
24. K. Robinson and P. F. Whelan, "Efficient morphological reconstruction: a downhill filter," *Pattern Recogn. Lett.* **25**(15), 1759-1767 (2004).
25. C. Lantuejoul and F. Maisonneuve, "Geodesic methods in image analysis," *Pattern Recogn.* **17**, 117-187 (1984).
26. U. Weidner, "Digital surface models for building extraction," *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, A. Gruen, ed., Birkhauser Verlag, Basel, 193-202 (1997).
27. B. Ameri, "Automatic Recognition and 3D Reconstruction of Buildings from Digital Imagery," PhD Thesis, Institute of Photogrammetry, Stuttgart University, DGK-C 526 (2000).
28. H. Arefi and M. Hahn, "A hierarchical procedure for segmentation and classification of airborne LIDAR images," *Proc. of the IEEE International Geoscience and Remote Sensing Symposium IGARSS* **7**, 4950-4953 (2005).

Hossein Arefi is a researcher at the German Aerospace Center (DLR), Remote Sensing Technology institute (IMF). He received his BS degree in surveying engineering from Tehran University in 1995 and MS degree in photogrammetry and geoinformatics from the Stuttgart University of Applied Sciences in 2002 and his PhD degree in photogrammetry and remote sensing from Bundeswehr University Munich in 2009. His current research interests include generating and analyzing digital terrain models produced from satellite stereo images as well as extraction and 3D reconstruction of the buildings from high resolution DSM data.

Pablo d'Angelo is a researcher at the Remote Sensing Technology Institute at the German Aerospace Center. He received his Diploma theses in computer engineering from the University of Applied Sciences Ulm in 2004 and his PhD degree from Bielefeld University in 2007. His current research interests include DSM generation from optical stereo images and camera calibration.

Helmut Mayer is professor of photogrammetry and remote sensing at Bundeswehr University Munich, Germany since 1999. He received his Diploma and PhD degrees in geodesy and geoinformation from Technische Universitaet Muenchen in 1990 and 1993, respectively. He is the author of more than 50 journals and refereed conference papers. His current research interests include highly precise direct 3D reconstruction from images from unmanned aircraft

systems (UAS) and the automatic extraction of topographic objects, particularly facades from terrestrial images. He is a member of the technical committee of the German pattern recognition association (DAGM).

Peter Reinartz is head of the Department of Photogrammetry and Image Analysis at the German Aerospace Centre (DLR), Remote Sensing Technology Institute (IMF). He also holds a professorship for geoinformatics at the University of Osnabrueck. After his Diploma in theoretical physics at the Ludwig-Maximilians-University of Munich he received his PhD (Dr.-Ing.) in civil engineering from the University of Hannover. He has more than 20 years of experience in image processing and remote sensing and over 150 publications in these fields. His main interests are in direct georeferencing, stereo-photogrammetry and data fusion of optical and radar data, generation of digital elevation models, and object based image analysis. He is member of IEEE.