

A NOVEL 3D CITY MODELLING APPROACH FOR SATELLITE STEREO DATA USING 3D ACTIVE SHAPE MODELS ON DSMS

Beril Sirmacek, Hannes Taubenboeck, Peter Reinartz

German Aerospace Center (DLR), Earth Observation Center (EOC)
PO Box 1116, 82234, Wessling, Germany
(Beril.Sirmacek, Hannes.Taubenboeck, Peter.Reinartz)@dlr.de

Commission III/4

KEY WORDS: Building, Modelling, City, DSM/DTM

ABSTRACT:

Since remote sensing provides new sensors and techniques to accumulate stereo data on urban regions, three-dimensional (3D) representation of these regions gained much interest for various applications. 3D urban region representation can e.g. be used for detailed urban monitoring, change and damage detection purposes. In order to obtain 3D representation, one of the easiest and cheapest way is to use Digital Surface Models (DSMs) which are generated from very high resolution stereo satellite images using stereovision techniques. Unfortunately after applying the DSM generation process, we cannot directly obtain a full 3D urban region representation. In the DSM which is generated using only one stereo image pair, generally noise, matching errors, and uncertainties on building wall locations are very high. These undesirable effects prevents a DSM to provide a realistic 3D city representation. Therefore, some automatic techniques should be applied to obtain real 3D city models using DSMs as input. In order to bring a solution to the existing problems in this field, herein we propose a fully automated approach based on the usage of a novel 3D active shape model. Our experimental results on DSMs of Munich city which are obtained from different satellite (Cartosat-1, Ikonos, WorldView-2) and airborne sensors (3K camera, HRSC, and LIDAR) indicate possible usage of the algorithm to obtain 3D city representation results automatically.

1 INTRODUCTION

As satellite and airborne sensor technology provides higher imaging qualities, especially 3D representation of cities gained much interest for various applications. For obtaining 3D representation, Digital Surface Models (DSMs) can be generated from optical stereo satellite or aerial images using stereovision techniques, or they can also be obtained by using LIDAR sensor technology. As a challenge, for satellite data, in most of the cases, just one stereo image pair is available for DSM generation. Unfortunately, after applying an automatic DSM generation process, due to the occlusion effects and stereo matching errors these DSMs do not correctly represent 3D city models with steep building walls and detailed rooftop representations. This is a major problem especially for DSMs which are generated over city centers, since many regions are occluded by dense and complex building structures. Although these occluded regions can be filled by interpolation, these techniques lead to a decrease in sharpness of building walls. Besides, deficiencies in the stereo matching process may cause noise within DSMs, e.g. due to shadow areas. Therefore, automatically obtaining 3D city models from DSMs is still an open and challenging problem for researchers.

In the previous work there is a wide variety of studies on building detection and shape extraction from two-dimensional single satellite or aerial images. The earliest studies in this field generally depend on edge and line extraction (Krishnamachari and Chellappa, 1996, Irvin and McKeown, 1989, Davis, 1982). Unfortunately, these methods generally fail to detect individual buildings which have highly textured rooftops or which appear in complex environments. In order to cope with this problem, Saeedi and Zwick (Saeedi and Zwick, 2008) combined edge information with graph based segmentation results of the region. Many researchers developed more advanced methods to extract shapes of the detected buildings (Karantzas and Paragios, 2009, Cui et al., 2008, Benedek et al., 2009). Sirmacek and Unsalan (Sirmacek and Unsalan, 2010) developed a fast method to detect shapes

of rectangular buildings which depends on growing a rectangular active shape. Unfortunately, they could not detect other building shapes with their approach. With the upcoming availability of DSMs from optical stereo images and from light detection and ranging measurements (LIDAR), many researchers began to pay attention to building detection from these data. Wurm et al. (Wurm et al., 2011), extracted 3D block models using an object-oriented approach based on data fusion from LIDAR and VHR optical imageries. Rottensteiner et al. (Rottensteiner et al., 2007) applied the DempsterShafer fusion of airborne laser scanner (ALS) point clouds and multispectral images for building detection. Haala et al. (Haala and Brenner, 1999) proposed a method to reconstruct building rooftops using surface normals extracted from LIDAR DSM data. They assumed that building boundaries are detected previously. In a following study (Haala et al., 1998), they detected building boundaries automatically by classifying a DSM and corresponding color image before applying their automatic rooftop reconstruction method. Brenner et al. (Brenner et al., 2001) discussed rooftop type derivation methods using given groundfloor shapes. They also projected two-dimensional terrestrial building facade images to generated models and obtain realistic 3D city representations. Brunn and Weidner (Brunn and Weidner, 1997) used surface normals on a DSM to discriminate buildings and vegetation. After detecting buildings, they measured the geometry of rooftops using surface normals and they interpolated polyhedral building descriptions to these structures. Fradkin et al. (Fradkin et al., 1999) proposed a segmentation based method to reconstruct 3D models of dense urban areas. To this end, they used very high resolution color aerial images and disparity maps. Canu et al. (Canu et al., 1996) used a high resolution DSM, which is obtained by stereo matching techniques, in order to reconstruct 3D buildings. First, they segmented the DSM into homogeneous regions. Then, they interpolated flat surfaces on these regions. Ortner et al. (Ortner et al., 2002) used a point process (Jacobsen, 2005) to model urban areas. They represented urban areas as interacting particles

where each particle stands for an urban object. Preknowledge about building shapes is used to model these particles. Arefi et al. (Arefi et al., 2008) extracted above-ground objects from LIDAR data. Then, 3D buildings are reconstructed by hierarchical fitting of minimum boundary rectangles (MBR) and a RANSAC based straight line fitting algorithm. Kada and McKinley (Kada and McKindley, 2009) introduced an approach for the automatic reconstruction of 3D building models. Again they used existing building groundfloor plans and LIDAR DSMs. Using building footprints they decomposed the building shape into sets of non-intersecting cells, and for each cell the rooftop is reconstructed by checking the normal directions of the DSM. Tournaire et al. (Tournaire et al., 2010), developed a stochastic geometry based on an algorithm to detect building footprints from DSM data which have less than 1m resolution. They tried to fit rectangles on the buildings using an energy function and prior knowledge about buildings. To minimize the energy function, they used a Reversible Jump Monte Carlo Markov Chain (RJCMC) sampler coupled with a simulated annealing algorithm which leads to an optimal configuration of objects. Maas (Maas, 1999) used maximum slope values in order to determine best fitting rooftop shapes to generate 3D building models. Valero et al. (Valero et al., 2008) developed a feature extraction and classification based method to classify building roofs into two classes as flat-roof and gable-roof. They estimated ridge-line positions which are based on skeletons of groundfloor plans. They provided the difference between the average roof outline height and the average ridge-line height as first feature, and the norm of the orthorectified image gradient as second feature for the support vector machine (SVM) classifier. In all introduced studies, good results are achieved generally using very high resolution (better than 1 m spatial resolution) DSMs which are generally generated from airborne images or LASER scan data. However, enhancement of buildings in low resolution urban DSM data which are generated from satellite images is still an open research problem. On the other hand, generally previous approaches require manual extraction of building outlines or providing groundfloor maps as input. In order to bring an automated solution to this problem, in previous work we have proposed a novel technique for obtaining 3D city representations by applying a building shape and rooftop-type detection approach to DSMs (Sirmacek et al., 2012). We started by applying local thresholding to raw DSMs in order to extract high urban objects which can indicate building locations. We have extracted building shapes from regions which are obtained from a thresholding result by using a binary active shape growing algorithm. This methodology depends on growing rectangular shapes in elongated segments which are detected in binary masks obtained by thresholding the DSM. After extracting the building shapes, we generated 3D models by understanding the building rooftop-types. Herein, we follow a similar approach to reconstruct 3D city models, however for active shape growing we propose a novel approach which uses 3D information in calculating shape fitting criteria. Using this new method, we increase the robustness of complex building shape extraction which in turn increases robustness of 3D reconstruction. Besides introducing a new methodology, our experiments also provide an insight on applicabilities of DSMs obtained from different sensors.

2 DETECTING POSSIBLE BUILDING SEGMENTS FROM DSMs

In this step, we would like to detect approximate building locations from the DSM before extracting building shapes. If a digital terrain model (DTM) of the region is available, we could use it to calculate a normalized digital elevation model (nDEM). In a nDEM, ground height is referenced to zero, therefore it only pro-

vides information about building heights independent from the height of the terrain. If a nDEM could be calculated, we could simply threshold it with a constant value in order to obtain high objects which can represent buildings or trees. In our study, we segment high objects directly from the DSMs by applying a local thresholding. Therefore, the algorithm can be also used for regions which do not have corresponding DTM data. In local thresholding, a 100×100 pixel size sliding window is used over the DSM, and a new threshold value is calculated for each region under the sliding window. This window size is chosen by considering approximate building sizes in given DSMs of the study region. However, the thresholding result does not differ significantly with slight changes of window size or with slight changes of input image resolution. Therefore, we can use the same window size for our input DSMs with different geometric resolutions.

After applying local thresholding to the DSM ($D(x, y)$), we obtain a binary image ($B_D(x, y)$) where high objects are labeled with value 1. We apply labeling to $B_D(x, y)$ to obtain its connected components (Sonka et al., 1999). Here each connected component represents a building segment. If the size of a connected component is less than R pixels we discard it since these small regions generally correspond to tree clusters. Considering geometric resolutions of input DSMs, we assume the R value as 100, since building objects cannot be smaller than this pixel size in our input DSMs. However, this value should be fixed by considering minimum sizes of the buildings in study regions before starting to run the algorithm on DSMs. In Fig. 1(a) and (b), we represent a subpart of the $D(x, y)$ and obtained $B_D(x, y)$ thresholding result respectively. Unfortunately, due to the low resolutions or surrounding trees around the building, thresholding result does not directly represent the building shape. However, it gives an idea about the approximate shape of the building.

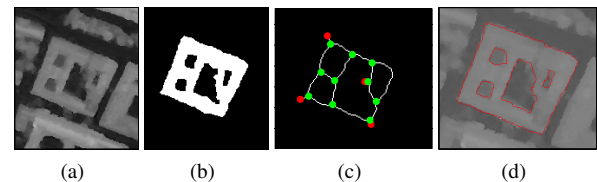


Figure 1: (a) A sub-part of the original Worldview2 satellite DSM ($D(x, y)$), (b) After applying local thresholding (sub-part of $B_D(x, y)$), (c) Skeleton of the building in the same sub-part of $B_D(x, y)$, (d) Detected building shape.

In the next step, we use the detected approximate segments to understand building complexity and to run our 3D active shape growing method.

3 EXTRACTING BUILDING SHAPES

In a previous study, Sirmacek and Unsalan (Sirmacek and Unsalan, 2010) proposed an automatic rectangular binary active shape growing approach (called box-fitting). First they used color invariant features to extract possible building rooftop segments. Mass centers of the rectangular segments are assumed as seed-points (as approximate building centers). Seed-point locations are used to grow a virtual active rectangular shape based on an energy criteria. In previous studies (Sirmacek et al., 2010) and (Sirmacek et al., 2012), we have used this binary active shape growing approach to detect complex building shapes from a binary $B_D(x, y)$ approximate building segment mask. First, we started by deciding if the building segment is complex or not. If there are inner yards (holes) inside of the segments, we assumed them as complex shape. We make this decision by computing an Euler number on binary building segment (Horn, 1986). If a

building segment is in a complex shape, we divided the building segment into elongated pieces using its skeleton. To do so, we detected junctions and endpoints of the building segment skeleton. We divided the skeleton into pieces by removing these junction pixels from the skeleton. For each obtained skeleton piece, we divided it again into l pixel length pieces if it is longer than l pixels. We assume center pixels of obtained skeleton pieces as our seed-point locations to run the box-fitting algorithm. Herein, similarly to our previous applications, we assume l value as equal to 20 pixels. In our previous study (Sirmacek et al., 2012), we provided effects of choosing l value to the shape detection result in detail.

For detecting complex building shapes, herein we follow a similar methodology. However instead of using the binary active shape growing approach in each seed-point location, we propose a novel active shape growing approach based on the usage of 3D information. To do so, after extracting (x_s, y_s) seed point locations as we describe in (Sirmacek et al., 2012) in detail, we start to grow our active rectangular shapes in each seed point location by regarding the height information. We assume that (x_v^n, y_v^n) array holds the pixel coordinates for n th edge of the virtual rectangular shape. Iteratively, we sweep each edge to the outwards growing direction if the edge pixels satisfy $(\max(D(x_v^n, y_v^n)) - \min(x_v^n, y_v^n) < \delta)$ inequality ($n \in [1, 2, 3, 4]$). Here, δ threshold is the minimum building height that we would like to detect in the region. In our application we assume δ as equal to 3, which means that we assume the buildings to be higher than 3 meters to be detected. When the growing process stops for each edge, we calculate the final energy value by using the equation that we represent in Eqn. 1. In the equation, $m(\cdot)$ represents the mean value. For the same seed-point, we apply growing process for all $\theta \in [0, \pi/6, \pi/3, \pi/2, 2\pi/3, \dots, 2\pi]$ angles with $\theta_{dif} = \pi/6$ radian turning steps. As we discussed in detail in (Sirmacek et al., 2012), by reducing θ_{dif} step sizes, we can obtain more accurate approximations, however in this case we need more computation time. After calculating E_θ for all θ angles, herein we pick the estimated box which exhibits the highest E_θ energy as detected building shape. Since most buildings appear like compositions of rectangular building segments, it makes sense to extract rectangular shapes on buildings. The main advantage of using the box-fitting approach is that approximate building shapes still can be found even if the building edges are not well-determined, or even if there are trees adjacent to the building facades. However, other region growing algorithms fail to extract an object shape in these cases, since the growing region can flow out easily when the parameters are not set precisely.

$$E_\theta = 4 \times m(B_D(x, y) \times D_i(x, y)) - \sum_{n=1}^4 m(D(x_v^n, y_v^n)) \quad (1)$$

For complex buildings, after fitting a chain of boxes, discontinuity between adjacent boxes should be smoothed. For this purpose, we simply benefit from morphological operations. First, we start with filling inside of the detected binary boxes with 1 value in $B_B(x, y)$ binary mask. Then, we apply morphological dilation and erosion operations respectively to the detected boxes, using a disk shaped structuring element with radius 1. After this operation small discontinuity between adjacent boxes can be smoothed. Final improved building shapes are kept in a new $B(x, y)$ binary mask.

The detected shape for our sample building can be seen in Fig. 1.(d). Obtained building shape is used to insert sharp building walls into our 3D model.

3.1 Rooftop-type Classification

After detecting building groundfloor shapes, we focus on reconstruction of rooftops. For this purpose, we benefit from our previous approach that we represent in (Sirmacek et al., 2012). We first start with ridge-line and tower detection. Using obtained ridge-lines we classify rooftops as flat or gable type. Obtained information is used to insert realistic models into our 3D city representations.

The ridge-line detection approach is based on derivative calculation over the DSM. We use the following derivative filter. For a symmetric Gaussian function $G(x, y) = \exp(-(x^2 + y^2))$, it is possible to define basis filters G_{p0} and $G_{p\frac{\pi}{2}}$ as follows,

$$G_{p0} = \frac{\partial}{\partial x} G(x, y) = -2x \exp(-(x^2 + y^2)) \quad (2)$$

$$G_{p\frac{\pi}{2}} = \frac{\partial}{\partial y} G(x, y) = -2y \exp(-(x^2 + y^2)) \quad (3)$$

We can find a derivative in an arbitrary θ direction using following filter function,

$$G_{p\theta} = \cos(\theta)G_{p0} + \sin(\theta)G_{p\frac{\pi}{2}} \quad (4)$$

We convolve our DSM with this derivative filter in $\theta \in [0, \pi/12, \dots, 23\pi/12]$ directions as follows,

$$J_\theta = D(x, y) * G_{p\theta} \quad (5)$$

If θ angle is perpendicular to the building ridge-line orientation, then one side of the building rooftop gives positive response, and the other side of the building rooftop gives negative response to the filter. Assuming $B^j(x, y)$ is the j th connected component in $B(x, y)$ binary building segment matrix, we assume that two sides of the building rooftop (RP_θ^j and RN_θ^j) can be extracted as follows,

$$RP_\theta^j = B^j(x, y) \times (J_\theta > 0) \quad (6)$$

$$RN_\theta^j = B^j(x, y) \times (J_\theta < 0) \quad (7)$$

Since we have no pre-information about building orientations, we should do the derivative filtering in all possible orientations. Therefore, we calculate $\sum_\theta RL_\theta^j$ for $\theta \in [0, \pi/12, \dots, 23\pi/12]$ directions. Building ridge-lines will be extracted in θ_j filtering angle which is almost perpendicular to the ridge-line orientation. However the ridge-line will be also detected in $\theta_j - \pi/12$ and $\theta_j + \pi/12$ neighbor filtering directions. Therefore, the ridge-line will have a value of higher than 2 in the $\sum_\theta RL_\theta^j$ result. As a result, the ridge-line of j th building rooftop can be obtained by calculating $R^j(x, y) = \sum_\theta RL_\theta^j > 2$ and eliminating connected components which are less than 10 pixels in order to eliminate redundant information coming from small objects on rooftop.

Towers around buildings are very important attributes to be considered in 3D model generation. Therefore, beside ridge-line detection and rooftop classification, tower detection from the input DSM is needed. To this end, in our study we detect maximum

values of the DSM (m_x) for each building segment. For j th building which is held as j th connected component in $B(x, y)$ binary building mask, the maximum height value is calculated as; $m_x = \max(B^j(x, y) \times D(x, y))$. Then, we threshold the DSM at building region using the following equation; $t^j(x, y) = (B^j(x, y) \times D(x, y)) > 0.8m_x$. Here, we obtained our threshold value by extensive tests. Finally, we apply size thresholding to $t^j(x, y)$ binary mask. In order to make the size thresholding process independent from the geometric resolutions of the input DSM, we compare the obtained region in $t^j(x, y)$ with the total sizes of the j th building segment. If the size of the region in $t^j(x, y)$ is smaller than 1/10 size of $B^j(x, y)$ segment, then $t^j(x, y)$ is assumed as a binary mask which holds a tower region inside. Otherwise, all pixel values in $t^j(x, y)$ matrix are changed with zero values. After applying the same methodology to all building segments in $B(x, y)$, union of all $t^j(x, y)$ matrices give us a binary mask which holds the tower regions. Binary tower matrix ($T(x, y)$) is obtained by calculating $T(x, y) = \bigcup t^j(x, y)$ for all j values.

Next, we use detected ridge-lines for roof-type classification and 3D rooftop model reconstruction purposes. We benefit from detected roof ridge-lines to classify rooftops as 'flat roof' or 'gable roof' type.

If there is no ridge-line detection result on a building segment, we assume that building as flat roof. If there is a ridge-line on the j th building segment, then we calculate mean of DSM height values on ridge-line location by calculating $\sum_{(x,y)} R^j(x, y) \times D(x, y) / M$, where M is the total number of ridge-line pixels in $R^j(x, y)$ binary matrix. We also calculate mean of DSM height values on building border by calculating $\sum_{(x,y)} B^j(x, y) \times D(x, y) / N$, where N holds the total number of building border pixels in $B^j(x, y)$ binary matrix. If the difference between these two mean values is lower than 2 meters, then we assume the rooftop as a *flat rooftop*. Otherwise, it is assumed as *gable rooftop*. This 2 meters criteria is obtained by observing *gable rooftop* characteristics over test area.

3.2 Generating 3D City Models

In raw DSM data, both building walls and rooftops need to be improved, since generally building walls make a soft transition from rooftop to the ground and the rooftop contains noise. In addition to that, raw DSM may include spurious values coming from stereo image matching errors in the DSM generation process. Therefore, in order to obtain better 3D representation, we should sharpen building walls and towers, and we should also enhance the rooftop view.

First, we start with generating a zero matrix $D_2(x, y)$ with the same size with $D(x, y)$ matrix. New height values belonging to objects in the city will be stored in $D_2(x, y)$ matrix. In order to eliminate noise in non-built regions, we apply median filtering to the original DSM ($D(x, y)$) by using a 3×3 pixel size window, and we obtain smoothed DSM ($D_f(x, y)$). For non-built areas, or in other saying for (x, y) coordinates which satisfy $B(x, y) = 0$, we apply $D_2(x, y) = D_f(x, y)$. That means; we assign smoothed ground height values for non-built regions. As building wall, we insert a single height value to each building boundary which is stored in $B(x, y)$ binary building shape matrix. For each building, the building height value is assigned by calculating the mean of $D(x, y)$ values on building boundary pixels. After smoothing the noise on the ground and inserting sharp building walls, finally rooftop height values are assigned to $D_2(x, y)$ matrix. Rooftop height assignment is done by considering the roof-type. If the roof is classified as a flat roof, then only

one single height value is assigned to all building areas which is equal to building wall heights. If the rooftop is classified as a gable-roof, we follow different approaches depending on the building complexity. If the building is not detected as a complex building as structure type (if the Euler number is equal to 1), then we can find polygons which define the rooftop in 3D space. To this end, we detect corners of the building segment using Harris corner detection algorithm (Harris and Stephens, 1988) over panchromatic image of the test region. Besides, we also detect endpoints of the building ridge-line. We pick each building corner one by one. Then, we find the closest ridge-line endpoint. A line between building corner and the closest ridge-line endpoint can divide the rooftop into polygons. A detailed demonstration of this approach is illustrated in (Sirmacek et al., 2012). Height values of rooftop polygons are assigned to corresponding pixels in $D_2(x, y)$ matrix. If the building is detected as a complex structure or if the building ridge-line could not be extracted properly, unfortunately we cannot use the same idea for building rooftop reconstruction and more advanced rooftop reconstruction approaches are needed. Herein we leave the building rooftop reconstruction at this level. For complex building rooftops and for rooftops for which we cannot extract the ridge-line properly, we only insert corresponding pixel values from $D_f(x, y)$ matrix. Finally, for tower regions which are stored in $T(x, y)$ binary matrix we insert the original DSM values (not filtered DSM values) to $D_2(x, y)$ matrix from $D(x, y)$.

4 EXPERIMENTS

4.1 Data Sets

The investigated data set consists of six DSMs which are obtained from different sensors covering each parts of the city of Munich, Germany. Three of them are obtained from the satellite sensors Cartosat-1, Ikonos, WorldView-2 and will show the potential of stereo satellite data to derive and enhance building models depending on their spatial resolution. All satellite data are along-track stereo data, since taking data from different dates is not appropriate due to radiometric changes and different shadow properties, which reduce the image matching density and quality significantly. The Cartosat-1 sensor on-board the Indian satellite IRS-P5 collects panchromatic data with two pushbroom cameras having a ground sampling distance (GSD) of approximately 2.5 m and a stereo angle of 31° . They are delivered with RPC values in a level called Ortho Ready, which is equivalent to Level 1A. RPCs have to be refined by ground control using an affine correction. The data set used has been acquired in May 2008. The whole area covered is about $30 \text{ km} \times 30 \text{ km}$, equivalent to one scene.

The Ikonos data used in this paper exhibit a GSD of 1 m and a stereo angle of approximately 10° , which is quite suitable for DSM generation in city areas (Reinartz et al., 2010). The data cover an area of approximately $10 \text{ km} \times 10 \text{ km}$ and have been acquired in July 2005; the processing level is "radiometric corrected", so also equivalent to Level 1A.

The WorldView-2 data used have a GSD of 0.5 m and are comprised of 2 stereo pairs acquired on the same day and orbit with an overlap in the test area, leading to 4 different viewing directions with stereo angles between 12° and 24° . These data have been acquired in July 2010 with the processing level "Level 2 Ortho Ready", which is equivalent to a projection on a plain for each single image. The overlapping for all 4 images is about $2 \text{ km} \times 10 \text{ km}$ in size. The DSMs for all three sensor data, generated by the above mentioned dense matching procedure (d'Angelo and

Reinartz, 2011), have the same sampling distance as the image data and cover all parts of the city center of Munich.

Two of the DSMs are obtained from airborne cameras, namely the High Resolution Stereo Camera (HRSC) and the 3K-camera system, both operated by DLR. The HRSC-AX (High Resolution Stereo Camera Airborne Extended) is a pushbroom camera and is especially designed to acquire stereo aerial photographs from 5 viewing directions. The image data exhibit a GSD of 1 m while the DSM, generated as described in (Hirschmueller et al., Springer, DAGM 2005, Wien, Austria).

The DLR 3K-camera system consists of three off the shelf digital frame cameras (Canon EOS 1Ds Mark, 16 MPixel) mounted on a platform, of which one is pointed nadir, the second pointed to the left and the third is pointed to the right according to the flight direction (Kurz et al., 2008). The data used are acquired from a flight altitude of approximately 1500 m, leading to a GSD of 25 cm. The DSM has a GSD of 50 cm and is derived using images with 80% overlap leading to 4 viewing directions per object. The airborne stereo data also cover parts of the Munich city center. Therefore for further investigations an area is used where DSM data from all sensors are available.

The sixth DSM, which is mainly used as a reference was acquired by an Airborne Laser Scanning (ALS) system and provided through the Bavarian Surveying Authorities, the 3D point cloud with a density of 2 points per square meter has been interpolated to a regular sampling distance grid of 1 m spacing.

4.2 Evaluation of Performances

In this section, we discuss the automatic 3D modelling performance of the proposed system on six different DSMs of a test building. In order to be able to calculate quantitative values for shape detection performance analysis, we use a binary building shape mask which includes groundfloor shapes of the buildings. This mask has been prepared by Munich municipality by using cadastral data. We apply pixel based performance analysis by counting the number of correctly detected building groundfloor pixels (*True Positives - TP*), and the number of false detected building groundfloor pixels (*False Alarm - FA*). By dividing the obtained numbers to the total number of building groundfloor pixels appearing in the groundtruth mask, in Table 1 we present *TP* and *FA* numbers as percentages. As can be seen in this table, the highest detection performance is obtained for the 3K, HRSC and LIDAR sensor DSMs, as expected since these are all derived from airborne data. But it has to be noted that also the WV2 DSM provides quite good results, with a quite low *FA* rate. On the other hand, the lowest detection performance is obtained on Cartosat1 sensor DSM since the low resolution of this sensor does not allow seeing neighbored building segments separately. The highest false alarm rate is obtained again on the Cartosat1 sensor DSM for the same reason. The lowest false alarm rate is obtained with LIDAR sensor DSM because of its high spatial resolution and since this DSM does not include trees which generally appear connected to the building facades.

In order to provide an insight on height estimation performances, we visualize a demonstration in Fig. 2. In Fig. 2.(a), we represent a profile of the sample building where we take height values. In 2.(b), the red dashed profile represents original WV2 DSM values, black dashed profile represents the LIDAR DSM values, while the blue continuous line represents automatically obtained reconstruction values by using WV2 DSM.

Finally, we compare the automatic reconstruction method that we propose herein with our previous reconstruction method which

is based on active shape growing using building edge information (Sirmacek et al., 2012). On the same building sample of WV2 DSM, our previous approach gives performance values as $TP = \% 65.02$ and $FA = \% 34.97$. Unfortunately, the previous approach cannot obtain successful detection results when the building edges cannot be detected correctly because of the noisy DSM or because of the objects on the building rooftops.

DSM Sensor	TP (%)	FA (%)
Cartosat1	14.34	98.56
Ikonos	78.96	22.00
WorldView-2	89.83	10.16
3K	94.01	6.62
HRSC	95.09	14.81
LIDAR	95.02	0.81

Table 1: Pixel Based Building Groundfloor Detection Performances for Six Different DSMs

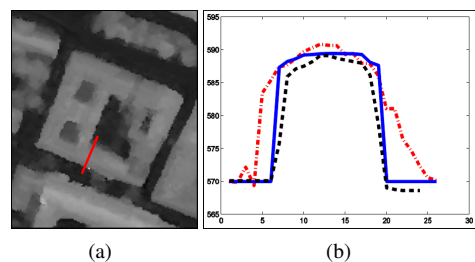


Figure 2: (a) The slice of the sample building where the height values are taken. (b) Red dashed profile represents original WV2 DSM values, black dashed profile represents the LIDAR DSM values, blue continuous line represents automatically obtained reconstruction values by using WV2 DSM.

5 CONCLUSIONS

Developing remote sensing technology and methods offer new and low-cost approaches such as DSM generation based on stereo satellite image matching principle. Herein, we introduced a novel method for automatic 3D detailed city modeling based on building shape, tower, and rooftop ridge-line extraction. Using the proposed approach we could generate 3D city models with high details even by using satellite images. Especially for regions which cannot be covered by airborne measurements, or for fast map updating or damage assessment purposes these data are well suited. Besides proposing a novel and robust approach for 3D city modeling, we provided a detailed assessment of the algorithm performance for different sensor data. For this purpose, we used DSMs which are obtained from different satellite (Cartosat-1, Ikonos, WorldView-2) and airborne sensors (3K camera, HRSC, and LIDAR). Beyond the development of the fine detailed 3D city models, we believe that the provided performance analysis over different sensor DSMs presents an important information about the capabilities of the different sensors and their remotely sensed stereo data.

REFERENCES

- Arefi, H., Engels, J., Hahn, M. and Mayer, H., 2008. Levels of detail in 3d building reconstruction from lidar data. In Proceedings of International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences 37, pp. 485–490.
- Benedek, C., Descombes, X. and Zerubia, J., 2009. Building extraction and change detection in multitemporal aerial and satellite images in a joint stochastic approach. INRIA, Paris, France, Tech. Rep.

- Brenner, C., Haala, N. and Fritsch, D., 2001. Towards fully automated 3d city model generation. Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images (III).
- Brunn, A. and Weidner, U., 1997. Extracting buildings from digital surface models. In Proceedings of International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences.
- Canu, D., Gambotto, J. and Sirat, J., 1996. Reconstruction of building from multiple high resolution images. In Proceedings of International Conference on Image Processing.
- Cui, S., Yan, Q. and Liu, Z., 2008. Right-angle building extraction based on graph-search algorithm. International Workshop on Earth Observation and Remote Sensing Applications, EORSA 2008 1, pp. 1–7.
- d'Angelo, P. and Reinartz, P., 2011. Semi-global matching results on the isprs stereo matching benchmark. ISPRS Hannover Workshop, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.
- Davis, L., 1982. Hierarchical generalized hough transforms and line-segment based generalized hough transforms. Pattern Recognition 15 (4), pp. 277–285.
- Fradkin, M., Roux, M., Maitre, H. and Leloglu, U., 1999. Surface reconstruction from multiple aerial images in dense urban areas. In Proceedings of IEEE Computer Vision and Pattern Recognition 2, pp. 262–26.
- Haala, N. and Brenner, C., 1999. Extraction of buildings and trees in urban environments. ISPRS Journal of Photogrammetry and Remote Sensing 54, pp. 130–137.
- Haala, N., Brenner, C. and Anders, K., 1998. 3d urban gis from laser altimeter and 2d map data. In Proceedings of International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences 32, pp. 339–346.
- Harris, C. and Stephens, M., 1988. A combined corner and edge detector. Proceedings of 4th Alvey Vis. Conf. pp. 147–151.
- Hirschmueller, H., Scholten, F. and Hirzinger, G., Springer, DAGM 2005, Wien, Austria. Stereo vision based reconstruction of huge urban areas from an airborne pushbroom camera (hrsc). Pattern Recognition, 27th DAGM Symposium pp. 58–66.
- Horn, B., 1986. Robot vision. New York, McGraw-Hill pp. 73–77.
- Irvin, R. and McKeown, D., 1989. Methods for exploiting the relationship between buildings and their shadows in aerial imagery. IEEE Transactions on systems, Man and Cybernetics 19 (6), pp. 1564–1575.
- Jacobsen, M., 2005. Point-process theory and applications: Marked point and piecewise deterministic processes (Probability and its applications). First edition edn, Birkhaeuser Boston.
- Kada, M. and McKindley, L., 2009. 3d building reconstruction from lidar based on a cell decomposition approach. Proceedings of ISPRS Workshop on Object Extraction for 3D City Models, Road Databases and Traffic Monitoring Concepts, Algorithms and Evaluation, CMRT'09, Paris, France.
- Karantzas, K. and Paragios, N., 2009. Recognition-driven two-dimensional competing priors toward automatic and accurate building detection. IEEE Transactions on Geoscience and Remote Sensing 47 (1), pp. 133–144.
- Krishnamachari, S. and Chellappa, R., 1996. Delineating buildings by grouping lines with mrfs. IEEE Transactions on Image Processing 5 (1), pp. 164–168.
- Kurz, F., Ebner, V., Rosenbaum, D., Thomas, U. and Reinartz, P., 2008. Near real time processing of dsm from airborne digital camera system for disaster monitoring. Congress of the ISPRS, Beijing, China, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.
- Maas, H., 1999. Fast determination of parametric house models from dense airborne laserscanner data. ISPRS Workshop on Mobile Mapping Technology, Bangkok, Thailand.
- Ortner, M., Descombes, X. and Zerubia, J., 2002. Building extraction from digital elevation models. INRIA Research Report.
- Reinartz, P., d'Angelo, P., Krauss, T., Poli, D., Jacobsen, K. and Büyüksalih, G., 2010. Benchmarking and quality analysis of dem generated from high and very high resolution optical stereo satellite data. Convergence in Geomatics, CGC, ISPRS. ISPRS Symposium Commission I, June 2010, Calgary, Kanada, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.
- Rottensteiner, F., Trinder, J., Clode, S. and Kubik, K., 2007. Building detection by fusion of airborne laser scanner data and multi-spectral images: performance evaluation and sensitivity analysis. ISPRS Journal of Photogrammetry and Remote Sensing 62 (2), pp. 135–149.
- Saeedi, P. and Zwick, H., 2008. Automatic building detection in aerial and satellite images. International Conference on Control, Automation, Robotics and Vision, ICARCV 2008, Hanoi, Vietnam 1, pp. 623–629.
- Sirmacek, B. and Unsalan, C., 2010. A probabilistic framework to detect buildings in aerial and satellite images. IEEE Transactions on Geoscience and Remote Sensing pp. –.
- Sirmacek, B., d'Angelo, P., Krauss, T. and Reinartz, P., 2010. Enhancing urban digital elevation models using automated computer vision techniques. ISPRS Commission VII Symposium, Vienna, Austria.
- Sirmacek, B., Taubenböck, H., Reinartz, P. and Ehlers, M., 2012. Performance evaluation for 3-d city model generation of six different dsms from air- and spaceborne sensors. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 1 (99), pp. 1939–1404.
- Sonka, M., Hlavac, V. and Boyle, R., 1999. Image processing, analysis and machine vision. PWS Publications, Pacific Grove, CA, second edn.
- Tournaire, O., Bredif, M., Boldo, D. and Durupt, M., 2010. An efficient stochastic approach for building footprint extraction from digital elevation models. ISPRS Journal of Photogrammetry and Remote Sensing 65 (4), pp. 317–327.
- Valero, S., Chanussot, J. and Gueguen, P., 2008. Classification of basic roof types based on vhr optical data and digital elevation model. IEEE International Geoscience and Remote Sensing Symposium IGARSS'08, Boston, USA.
- Wurm, M., Taubenböck, H., Schardt, M., Esch, T. and Dech, S., 2011. Object-based image information fusion using multisensor earth observation data over urban areas. International Journal of Image and Data Fusion 2 (2), pp. 121–147.