3D-INFORMATION FUSION FROM VERY HIGH RESOLUTION SATELLITE SENSORS

Thomas Krauß,[*] Pablo d'Angelo, Georg Kuschk, Jiaojiao Tian, Tahmineh Partovi

German Aerospace Center (DLR), Earth Observation Center, Remote Sensing Technology Institute,
Oberpfaffenhofen, 82234 Weßling, Germany – Thomas.Krauss@dlr.de

KEY WORDS: Stereo satellites, DSM, 3D-objects, DTM, 3D-change-detection

ABSTRACT:

In this paper we show the pre-processing and potential for environmental applications of very high resolution (VHR) satellite stereo imagery like these from WorldView-2 or Pléiades with ground sampling distances (GSD) of half a metre to a metre. To process such data first a dense digital surface model (DSM) has to be generated. Afterwards from this a digital terrain model (DTM) representing the ground and a so called normalized digital elevation model (nDEM) representing off-ground objects are derived. Combining these elevation based data with a spectral classification allows detection and extraction of objects from the satellite scenes. Beside the object extraction also the DSM and DTM can directly be used for simulation and monitoring of environmental issues. Examples are the simulation of floodings, building-volume and people estimation, simulation of noise from roads, wave-propagation for cell-phones, wind and light for estimating renewable energy sources, 3D change detection, earthquake preparedness and crisis relief, urban development and sprawl of informal settlements and much more. Also outside of urban areas volume information brings literally a new dimension to earth oberservation tasks like the volume estimations of forests and illegal logging, volume of (illegal) open pit mining activities, estimation of flooding or tsunami risks, dike planning, etc. In this paper we present the preprocessing from the original level-1 satellite data to digital surface models (DSMs), corresponding VHR ortho images and derived digital terrain models (DTMs). From these components we present how a monitoring and decision fusion based 3D change detection can be realized by using different acquisitions. The results are analyzed and assessed to derive quality parameters for the presented method. Finally the usability of 3D information fusion from VHR satellite imagery is discussed and evaluated.

1. INTRODUCTION

Very high resolution (VHR) satellite missions like WorldView-2 or Pléiades allow the acqusition of (multi-)stereo imagery with ground sampling distances (GSD) of about half a metre. Sophisticated algorithms for dense image matching from such imagery allow the reconstruction of a detailed 3D point cloud from the acquired images with a vertical accuracy also in the range of one GSD.

Using the height-information together with the multi-spectral image data allows the classification, object-detection and extraction on a much higher level than from multispectral imagery alone. Since 3D data from optical imagery are always digital surface models (DSMs) this information is mostly useful especially in urban areas. In such areas the fusion of height- and spectral classification allows the potential detection of single objects like buildings, trees, roads and bridges and subsequently a more robust object based change detection in contrast to mostly used simple pixel based change detection algorithms. But also outside of cities the third dimension increases the worth of remote sensing data. Particularly height sensitive areas like mining or forest areas can be monitored more detailed. So the output of a mine can easily be calculated by measuring the volume differences of the mining pit and the waste dump as we show later. For deep mining or other undearearth activity also the volume change of detected waste dumps can be very valuable.

In this paper we describe the preprocessing steps necessary for generating elevation and classification data which can be used in turn for object detection, change detection and monitoring. For this first a DSM has to be derived from the (multi-)stero data. The resulting 3D data have to be fused with the original imagery to reduce errors and outliers. Afterwards the digital terrain model

---

[*]Corresponding author

(DTM) representing the ground level and the normalized digital elevation model (nDEM) containing the height of all elevated objects above the DTM are derived.

In parallel some multi-spectral classifications are executed, espacially the detection of vegetation, water and shadows. Afterwards all available data are fused for different purposes. If more scenes are available also all height- and spectral classified data can be fused for detailed and stable change detection and monitoring applications.

After an overview of the methods used we show in a few examples the monitoring from GeoEye and WorldView-2 data and the application of such data for change detection in an open pit mining area. In a second experiment we show 3D building detection in Ikonos summer and winter scenes and in a third experiment we use panchromatic Cartosat-1 stereo data showing a change detection application in forest areas. Finally a short conclusion on using fused multi-spectral VHR stereo data on monitoring applications from space closes the paper.

1.1 Preliminary work

This work is based on many algorithms and methods already published in other places. So the method used for dense stereo matching called Semi global matching was already described in principle in Hirschmüller (2005) and refined in dAngelo et al. (2008) to generates DSMs from space-borne data. The extraction of DTMs from such DSMs uses approaches as described in Arefi et al. (2009) or Krauß et al. (2011). The fuzzy-classification based on derived top-of-atmopsphere (TOA) ortho images is described in Krauß et al. (2012). An additional method used here is the shadow detection shown by Makarau et al. (2011). The basic approach for change detection used in this work is the robust change detection method described in Tian and Reinartz (2011) and Tian et al. (2014).

2. PRE-PROCESSING METHODS

2.1 DSM generation

The VHR sensors used here have already a very good absolute image positioning of e.g. about 4 m for WorldView-2 only requiring a slight shift for absolute geocorrection. But dense stereo matching requires a relative correlation between the stereo images of less than a pixel. To achieve this first good correlation points are derived automatically from the two stereo images and the shifts are applied to the sensor model of the images.

For reconstructing the 3D scene, we use dense stereo matching to make use of all the image information and compute a depth for each single image pixel. As for each image pixel multiple depth hypotheses need to be tested, a computational efficient cost function is needed for the image matching, which additionally has to be robust to potentially strong radiometric and perspective changes between the images. As the pure data term of the cost function is still prone to noise and performs poorly over large disparity ranges and homogeneously textured regions, a simple pixel-wise Winner-takes-all strategy solely based on the image matching function does not yield acceptable results, which is why we add additional smoothness constraints and solve the resulting optimization problem.

The Census transform CT (Zabih and Woodfill (1994)) encodes the local image structure within a small patch around a given pixel. It is defined as an ordered set of comparisons of intensity differences and therefore invariant to monotonic transformations which preserve the local pixel intensity order:

$$\xi(I(\mathbf{x}), I(\mathbf{x}')) = \begin{cases} 1 & \text{if } I(\mathbf{x}') < I(\mathbf{x}) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$CT(I, \mathbf{x}) = \bigotimes_{[i,j] \in D} \xi(I(\mathbf{x}), I(\mathbf{x} + [i,j])) \ , \quad (2)$$

with the concatenation operator $\bigotimes$ and an ordered set of displacements $D \subset \mathbb{R}^2$, which is chosen to be a 7×9 window due to its simple fitting into 64 Bit variables. The matching cost of different Census vectors s_1, s_2 is then computed as their Hamming distance $d_H(s_1, s_2)$ – number of differing bits – where highest matching quality is achieved for minimal Hamming distance, and the costs are normed to the real-valued interval $[0, 1]$

$$cost(\mathbf{x}, \gamma) = \frac{d_H\left(CT(I_1, \mathbf{x}), \ CT(I_2, F_{(1,2)}(\mathbf{x}, \gamma)) \right)}{\max_{i,j}\{d_H(s_i, s_j)\}} \quad (3)$$

Using such a 7×9 support window for the Census transform increases the robustness of the matching function against mismatches, especially when searching through a large disparity range as is very common in remote sensing data.

In the resulting disparity space image (DSI) we now search for a functional $u(\mathbf{x})$ (the disparity- or height-map), which minimizes the energy function arising from the matching costs (called data term E_{data}), plus additional regularization terms. As the data term is prone to errors due to some incorrect and noisy measurements, one often needs some smoothness constraints E_{smooth}, forcing the surface of the disparity map to be locally smooth.

$$u(\mathbf{x}) = \underset{u}{\arg\min} \left\{ \int_\Omega E_{data} + E_{smooth} \ d\mathbf{x} \right\} \quad (4)$$
$$= \underset{u}{\arg\min} \left\{ \int_\Omega C(\mathbf{x}, u(\mathbf{x})) + \nabla u(\mathbf{x}) \ d\mathbf{x} \right\}$$

This energy is non-trivial to solve, since the smoothness constraints are based on gradients of the disparity map and therefore cannot be optimized pixelwise anymore. Because of its simple implementation, low computational complexity and good regularization quality, we use semi-global matching (SGM) for optimization (Hirschmüller, 2005), minimizing the following discrete energy term

$$E(u) = \sum_{\mathbf{x}} C(\mathbf{x}, u(\mathbf{x})) +$$
$$\sum_{\mathbf{p} \in N_{\mathbf{x}}} P_1 \cdot T[|u(\mathbf{x}) - u(\mathbf{p})| = 1] +$$
$$\sum_{\mathbf{p} \in N_{\mathbf{x}}} P_2 \cdot T[|u(\mathbf{x}) - u(\mathbf{p})| > 1] \quad (5)$$

The first term is the matching cost, the second and third term penalties for small and large disparity discontinuities between neighboring pixels $\mathbf{p} \in N_{\mathbf{x}}$. The key idea is now not to solve the intractable global 2D problem, but to approximate it by combining 1D solutions from different directions $\mathbf{r}$, solved via dynamic programming

$$u(\mathbf{x}) = \underset{u}{\arg\min} \left\{ \sum_{\mathbf{r}} L_{\mathbf{r}}(\mathbf{x}, u(\mathbf{x})) \right\} \quad (6)$$

The aggregated cost $L_{\mathbf{r}}(\mathbf{x}, u(\mathbf{x}))$ in Equation (6) represents the cost of pixel $\mathbf{x}$ with disparity $d = u(\mathbf{x})$ along the direction $\mathbf{r}$ and is being computed as

$$L_{\mathbf{r}}(\mathbf{x}, d) = C(\mathbf{x}, d) \ + \ \min(L_{\mathbf{r}}(\mathbf{x} - \mathbf{r}, d),$$
$$L_{\mathbf{r}}(\mathbf{x} - \mathbf{r}, d - 1) + P_1 \ , \ L_{\mathbf{r}}(\mathbf{x} - \mathbf{r}, d + 1) + P_1 \ ,$$
$$\min_k L_{\mathbf{r}}(\mathbf{x} - \mathbf{r}, k) + P_2) \ - \ \min_k L_{\mathbf{r}}(\mathbf{x} - \mathbf{r}, k)$$
$$(7)$$

The penalties are set to default values $P_1 = 0.4$ and $P_2 = 0.8$ (with the raw costs normed to $[0, 1]$). For deeper insights the reader is referred to the paper of Hirschmüller (2005).

2.2 DSM correction and Ortho projection

The such derived disparity map d now fits exactly on the first image I. Missing values due to occlusion or mismatches can be filled using the spectral similarity of neighbouring pixels in the pan-sharpened first image I_{ps} with valid disparities as shown in fig. 1. Afterwards using the sensor-model – in the cases of VHR satellites mostly rational polynomial coefficients (RPCs) – a height map h can be calculated by converting the disparities d of each pixel to true ellipsoidal heights.

Applying the filled height map h to the original imagery I_{ps} using the sensor model delivers the ortho images I_o and applying h to itself the correctly geocoded DSM D as shown in figs. 2 and 3).
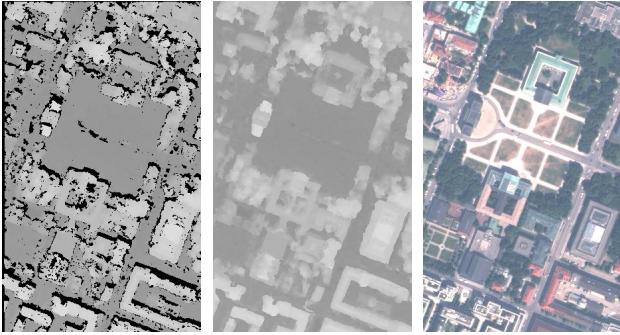
Figure 1: Height map fitting on first pan image, corrected height map, pan-sharpened first image (l.t.r.), example from a WorldView-2 scene of Munich, Königsplatz, 300 m × 500 m each

Figure 2: Ortho image of a GeoEye scene, 16 × 9 km

2.3 DTM extraction and classification

Based on the DSM D the terrain model (DTM) T representing the ground can be derived using the morphological method described in Krauß et al. (2011) based on the unfilled DSM. The difference of the surface and the terrain model gives the normalized DEM $N = D - T$. Combining the DSM D and DTM T allows such the extraction of a high objects mask $m_h = 1$ if $N > t_h$ and $m_h = 0$ else with a height-threshold $t_h \approx 3$ m.

In addition to the height classification also a spectral classification is needed for detection of urban objects or the change detection methods like the one described in Tian and Reinartz (2011). For this the multispectral imagery is converted first to a top of atmosphere (TOA) reflectances image I_{toa}. Afterwards a shadow mask m_s is calculated following Makarau et al. (2011) using the blue, green and red channel of I_{toa}. Applying the spectral fuzzy classification method described in Krauß et al. (2012) to I_{toa} allows the extraction of land cover probability maps. These maps contain the normalized digital vegetation index (NDVI) F_n and a bare soil (F_b) and water map F_w. From these the vegetation, bare soil and water mask m_v, m_b and m_w are derived using $m_v = 1$ if $F_n > t_v$, $m_b = 1$ if $F_b > t_b$ and $m_w = 1$ if $F_w < -t_w$ with the thresholds $t_v \approx 0.3$, $t_b \approx 0.5$ and $t_w \approx 0.1$.

Combining all classes gives a class map as shown in fig. 4. Using these classes the changes can be described as shown in tab. 1.

In mining areas the classification becomes more complex since the DTM does not follow the DSM correctly, so the mining area can show up any of the height changes. In this case only the real height change is of interest. Also the phenology of the trees have a great influence on the classification. In the WorldView-2 image taken in March most of the forest areas do not show up as high vegetation since the DSM generation process delivers mainly the more dominant ground. But such information can be used in monitoring purposes for tree type classification.
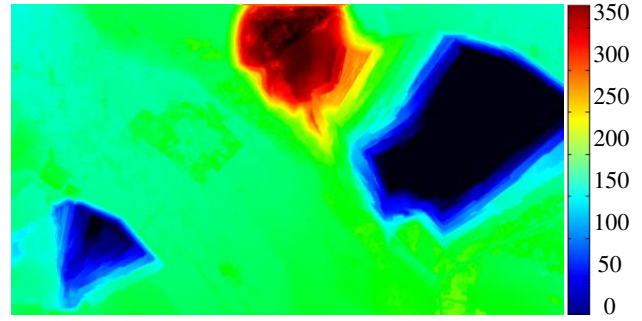

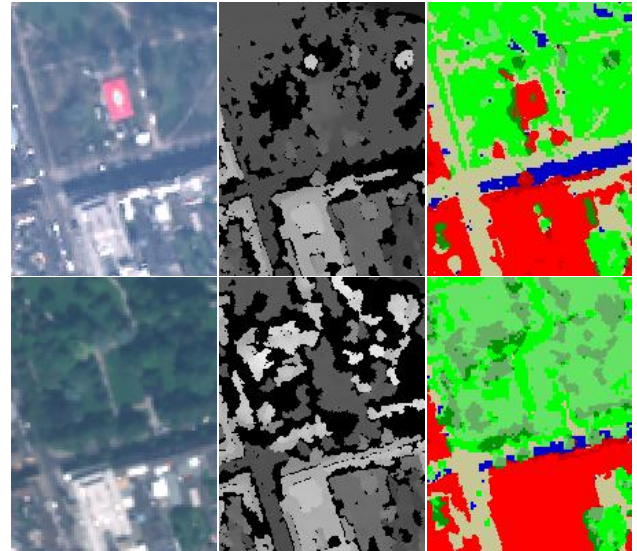Figure 3: DSM of the same GeoEye scene, 16 × 9 km, 0 to 350 m height

Figure 4: Ortho image, unfilled DSM and class map of the WorldView-2 (2011, top) and the GeoEye (2012, bottom) scenes (section 150 m × 200 m); Classes: red/dark red: high non-vegetation objects (mainly buildings)/with shadows, tan/blue: low non-vegetation objects (mainly bare soil)/with shadows, dark green/olive: high vegetation objects (e.g. trees)/with shadows, light green/olive: low vegetation objects (grassland)/with shadows

2.4 Change detection

Knowing all heights and class maps for scenes of two or more epochs allow the application of the robust difference method as described in Tian et al. (2014) for detecting height changes. The robust difference between the initial DSM D_1 and the second DSM D_2 for the pixel (x, y) can be defined as the minimum of differences computed between the pixel $D_2(x, y)$ in the second DSM and a certain neighborhood (with window size $2w + 1$) of the pixel $D_1(x, y)$ in the first DSM D_1. The robust positive and negative differences $X_{Pdif}(x, y)$ and $X_{Ndif}(x, y)$ relative to the pixel (x, y) are defined in equations 8 and 9, respectively:

$$X_{Pdif}(x, y) = \min_{x', y'} \left(\left(D_2(x, y) - D_1(x', y') \right), \right.$$
$$\left. \left(D_2(x, y) - D_1(x', y') \right) > 0 \right) \quad (8)$$

$$X_{Ndif}(x, y) = \max_{x', y'} \left(\left(D_2(x, y) - D_1(x', y') \right), \right.$$
$$\left. \left(D_2(x, y) - D_1(x', y') \right) < 0 \right) \quad (9)$$

with $x' \in [x - w, x + w]$ and $y' \in [y - w, y + w]$ in a small

Table 1: Possible changes depending on classified areas, empty fields mean "no change", v and $\bar{v}$ denote vegetation or non-vegetation, bld. stands for building

	low → low	low → high	high → low	high → high
$v \to v$		trees growed	trees logged	
$v \to \bar{v}$	harvested	bld. erected	trees logged	trees → bld.
$\bar{v} \to v$	planted	trees growed	bld. demolished	bld. → trees
$\bar{v} \to \bar{v}$		bld. erected	bld. demolished	

window around (x, y).

Eqs. 8 and 9 say that only the minimum value (greater than zero) in case of a positive change, or the maximum value in case of a negative change is taken, all within the defined window size. Typically window sizes used are 3×3 pixels up to 7×7 pixels depending on the DSM qualities and the difference in resolution between the two available DSMs.

To eliminate noise these robust height changes are filtered afterwards using morphological closing and opening operations. Also for a more accurate change detection of buildings the changes are masked using the shadow mask and other land cover masks as described above. Combining the classification results (vegetation /non-vegetation and high/low masks) with the detected changes allow for a more detailed classification of the changes.

3. EXPERIMENTS

3.1 WorldView-2 – GeoEye

In this experiment we use two VHR stereo scenes acquired on 2011-03-24, 10:58 (WorldView-2) and 2012-08-12, 10:39 (Geo-Eye) over Jülich. This area contains the city of Jülich, Germany and also the largest open pit mining area for lignite in Germany. Based on these stereo images first the DSMs D_1 and D_2 were derived as described in 2.1. Afterwards the DTMs T_1 and T_2 together with the nDEMs N_1 and N_2 and the pansharpened TOA ortho images $I_{toa,j}$ and all masks $m_{i,j}$ ($i \in \{h, s, v, b, w\}$, $j \in \{1, 2\}$) as described in 2.3 are calculated. Fig. 2 showed already the ortho image and fig. 3 the DSM of the GeoEye scene of this area.

Calculating the robust difference and the appropriate morphological filtering gives the changes as shown in fig. 5.
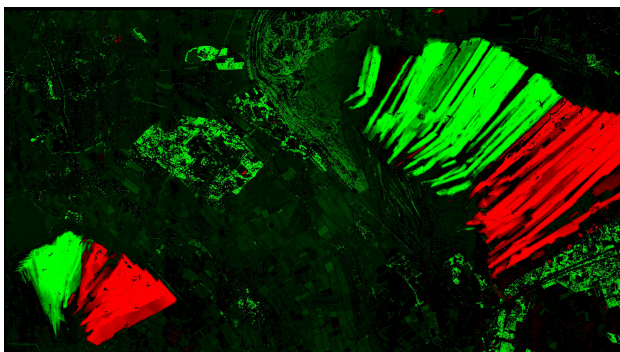

Figure 5: Volume changes between the WorldView-2 and GeoEye scene, 16 km × 9 km, red: negative changes, green: positive changes (0–60 m)

For a more detailed view fig. 6 shows a small (1100 m × 450 m) section with a newly erected building in the left top edge (in green) and a cut down forest in the right bottom (in red). The

types of objects (building, forest) and type of change (erection and logging) is hereby derived from the class-map and following tab. 1. All other green objects in this image are trees not visible in the March WV-DSM but detected in the August GE-DSM.
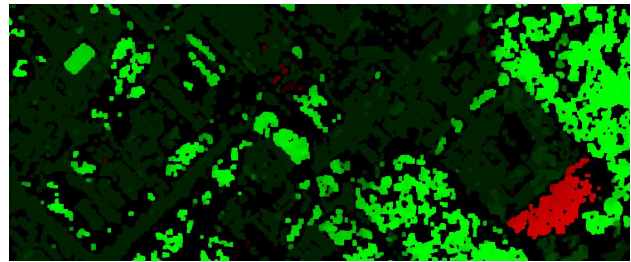

Figure 6: Volume changes between the WorldView-2 and GeoEye scene, small section 1100 m × 450 m, top left: newly erected building, bottom right: cut down forest

In the left bottom edge of fig. 5 the open pit mining area "Braunkohletagebau Inden" is visible. A positive or negative height change from any kind of vegetation to non-vegetation allows us to calculate the mined volumes. For this the positive (green) and negative (red) volumes of fig. 5 are calculated as 79.87 and 122.76 millions of cubic metres. The negative changes are the whole digged out volume containing the lignite and also the mine wastes. The positive changes are however only the dumped mine wastes. The mine wastes consist only of earth and stone with a density of $\rho_{soil} = 1500$ kg/m^3 while the lignite has a density of $\rho_{lignite} = 1000$ kg/m^3. This allows us to convert the volume changes to mass changes as shown in tab. 2.

Table 2: Validation of 3D change detection results for the opencast mining Inden (left bottom mining area); reference data from http://de.wikipedia.org/wiki/Tagebau_Inden

Value	Mine wastes	Digged lignite
Reference amount per year	80–85 Mio t/a	20–25 Mio t/a
Amount between images (506 days)	111–118 Mio t	28–35 Mio t
Volumes (m/ρ)	74–79 Mio m^3	28–35 Mio m^3
Measured volume changes	80 Mio m^3	43 Mio m^3

The two scenes were acquired with a distances of 506 days. So in tab. 2 first the reference amounts per year have to be scaled to this time distance. Afterwards the amount is converted to volumes and compared to the measured volumes from the 3D change detection between the two scenes. Finally we see that the measured volume changes from the described automatic change detection method comply very well with the amounts Wikipedia claims without any reference for this mine.

3.2 Ikonos building detection in summer and winter scenes

Our second example shows the building change-detection in different seasons of two Ikonos scenes from Dong-An (North-Korea) acquired 2006 and 2010 (snow covered) as shown in fig. 7 together with the derived DSMs in fig. 8.

By applying the change detection method described in 2.4 masking out only buildings following 2.3 a change probility map shown in fig. 10, left can be derived. Cross checking the results with a manually created reference change mask and thresholding the changes allows for a evaluation of the results (see fig. 9 for explanation of the colors). The overall accuracy for this example is 0.9920 while the kappa index of agreement is 0.7502. All qualitiy parameters of this evaluation are calculated pixel wise: true positives (TP): green, pixels found by algorithm and also in reference

Figure 7: Ortho images of Ikonos scenes of 2006 (left) 2010 (right, covered with snow), both sections 1000 m × 1000 m
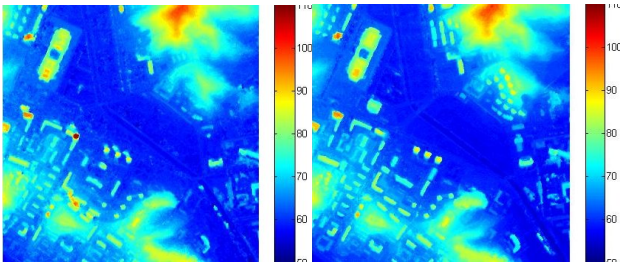

Figure 8: DSMs from the scenes of 2006 (left) 2010 (right), both sections 1000 m × 1000 m, color coded heights in m

mask; true negatives (TN): black, pixels detected unchanged and also unchanged in reference; false positives (FP): red, unchanged pixels found erroneously as changes by the algorithm; false negatives (FN): blue, changed pixels wrongly classified as unchanged; overall accuracy (OA) and the kappa index of agreement κ calculated following Congalton (1991):

$$OA = \frac{TP + TN}{N} \quad \text{and} \quad \kappa = \frac{OA - PA}{1 - PA}$$

with $N = TP + TN + FP + FN$ as the total number of pixels and PA as the hypothetical probability of agreement:

$$PA = \frac{(TP + FP) \cdot (TP + FN)}{N^2} + \frac{(TN + FP) \cdot (TN + FN)}{N^2}$$

Reference data Change Mask	Change	No-Change
Change	True detected	False alarm
No-Change	Missed alarm	

Figure 9: Color coding scheme of changes shown in change masks

3.3 Cartosat – Forest

The third experiment is based on pan-chromatic Cartosat-1 (IRS-P5) stereo scenes from a forest region near Oberammergau, Germany, acquired 2008 and 2009. These images are acquired with two stereo cameras (forward: $26°$, backward: $-5°$) with a resolution of 2.5 m. Two ortho images are shown in fig. 11. The DSMs created from the stereo imagery can be seen in fig. 12.

Applying a mean-shift over-segmentation to the ortho imagery and a region merging on the combination results in a land cover segmentation using only one pan-chromatic band (Tian et al.,
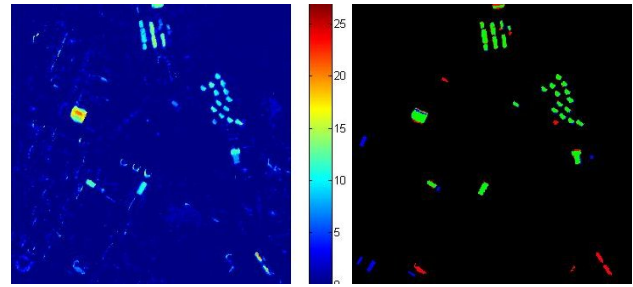

Figure 10: Detected changes in the Ikonos stereo DEMs of Dong-An, left: change probability (in %), right: change mask (green: true detected changes, blue: missed changes, red: erroneously detected changes)
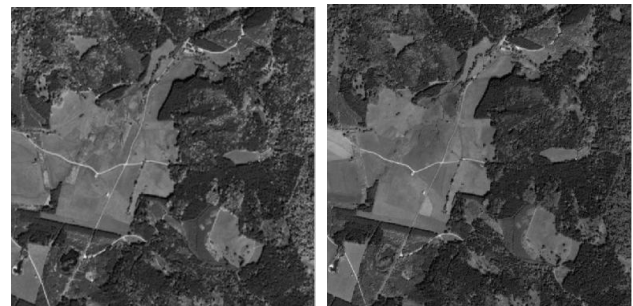

Figure 11: Cartosat ortho images of a forest area near Oberammergau 2008 (left) and 2009 (right)

2013). To analyze the changes the height changes from the DSMs and robust multi-level change features from the images are used. Putting this all together results in a change probability map shown in fig. 13, left.

For the change detection between the two Cartosat images the overall accuracy is 0.9945 and the kappa index 0.5879.

4. RESULTS AND DISCUSSION

Fusing height information with multispectral classification allows more detailed monitoring and better change detection. The results from the above experiments show very promising results of the proposed processing chain. The mining volumes are derived with an accuracy of about 4.5 % for the waste and 36 % for the digged coal. Also the accuracies calculated against manually derived change masks of the multi-season Ikonos building changes and the forest changes are very promising with accuracy values larger than 99 %.

The main problems found can be traced back to mismatches in the DSM generation process and problems in the spectral classification – especially in water and shadow areas. The outliers in the unfilled DSM propagate in the DSM filling process often into larger areas destroying the height mask for buildings. More sophisticated DSM correction methods have to be investigated to create more stable and reliable DSMs. Also better DSMs can be generated using three or more stereo images. First experiments with multi-stereo WorldView-2 and Pléiades imagery show that a third image improves the quality of the DSM drastically while a fourth or fifth image lead only to minor further improvements. Also scenes acquired in the same season lead to a more consistent classification for the change detection.
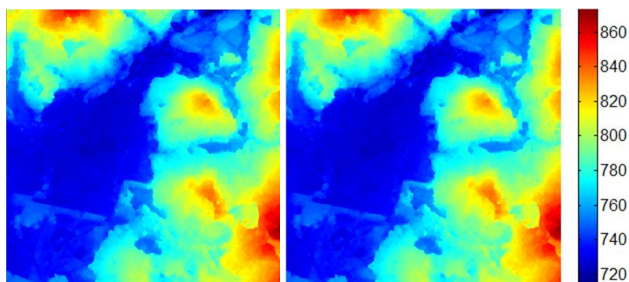
Figure 12: DSMs created from the Cartosat stereo images of a forest area near Oberammergau 2008 (left) and 2009 (right), heights in [m]
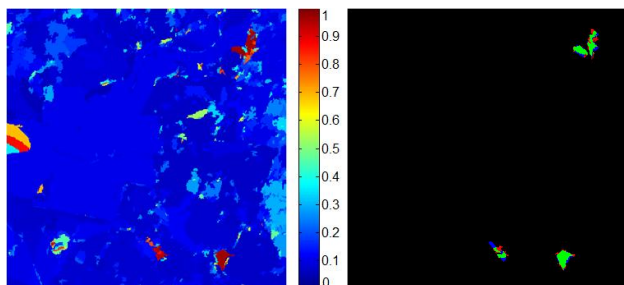

Figure 13: Resulting change map and masks generated in the Oberammergau test area, left:Region-based change probability map (0–1=0–100 %), right: Refined mask overlaid with reference data, green: true detected, red: false alarm, blue: missed alarm

5. CONCLUSION AND OUTLOOK

The presented methods for preparing and pre-processing very high resolution (VHR) satellite data open new possibilities on the remote sensing of the environment. Based on VHR satellite stereo imagery detailed 3D representations of urban areas or other environments can be prepared and used for simulation or monitoring issues. Besides the extraction of 3D objects also the much more precise change detection is an additional potential in monitoring the environment.

Using the method allows a potentially fully automatic change detection which is much more sophisticated than only image based methods as can be seen in the Ikonos summer/winter example. Also the monitoring of mining activities, illegal timber logging or monitoring the evolution of large cities are a good application area for the presented method.

But nevertheless the main disadvantage of the method is the need of pre- and post-elevation information. In most cases like disasters no pre-disaster stereo imagery or DSMs of the required resolution and region exists.

Unfortunately the Sentinel missions do not support stereo acquisitions. So we propose an additional sentinel mission covering the whole earth permanently like the indian Cartosat-1 (IRS-P5) mission with a two- or even better a three-line-stereo-scanner (like ALOS-Prism) at a ground sampling distance of the stereo scanners of at least 2.5 m and also a nadir looking multispectral VNIR scanner with also about 2.5 m GSD or better.

References

Arefi, H., dAngelo, P., Mayer, H. and Reinartz, P., 2009. Automatic generation of digital terrain models from cartosat-1 stereo images. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.

Congalton, R. G., 1991. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. Remote sensing of Environment 37(1), pp. 35–46.

dAngelo, P., Lehner, M., Krauß, T., Hoja, D. and Reinartz, P., 2008. Towards automated dem generation from high resolution stereo satellite images. In: IAPRS, Vol. 37, pp. 1137–1142.

Hirschmüller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, pp. 807–814.

Krauß, T., Arefi, H. and Reinartz, P., 2011. Evaluation of selected methods for extracting digital terrain models from satellite born digital surface models in urban areas. In: International Conference on Sensors and Models in Photogrammetry and Remote Sensing (SMPR 2011).

Krauß, T., Sirmacek, B., Arefi, H. and Reinartz, P., 2012. Fusing stereo and multispectral data from WorldView-2 for urban modeling. In: Proc. SPIE 8390, Vol. 83901X, p. 2012.

Makarau, A., Richter, R., Müller, R. and Reinartz, P., 2011. Adaptive shadow detection using a blackbody radiator model. IEEE Trans.Geosci. Remote Sens. 49(6), pp. 2049–2059.

Tian, J. and Reinartz, P., 2011. Multitemporal 3D Change Detection in Urban Areas Using Stereo Information from Different Sensors. In: International Symposium on Image and Data Fusion, Vol. 2011, pp. 1–4.

Tian, J., Cui, S. and Reinartz, P., 2014. Building change detection based on satellite stereo imagery and digital surface models. IEEE Trans. Geosci. Remote Sens. 52(1), pp. 406–417.

Tian, J., Reinartz, P., dAngelo P. and Ehlers, M., 2013. Region-based automatic building and forest change detection on Cartosat-1 stereo imagery. ISPRS Journal on Photogrammetry and Remote Sensing 79, pp. 226–239.

Zabih, R. and Woodfill, J., 1994. Non-parametric Local Transforms for Computing Visual Correspondence. Computer Vision - ECCV 1994 pp. 151–158. census transform, rank transform.