# Multi-spectral and Digital Elevation Model Information Fusion for Automatic 3D Change Detection

Houda Chaabouni-Chouayakh, Isabel Rodes Arnau and Peter Reinartz

*Remote Sensing Technology Institute (IMF)*

*German Aerospace Center (DLR)*

*Oberpfaffenhofen, Germany*

Houda Chaabouni-Chouayakh

German Aerospace Center (DLR)

Oberpfaffenhofen

82234 Wessling

Germany

houda.chaabouni@dlr.de

# Multi-spectral and Digital Elevation Model Information Fusion for Automatic 3D Change Detection

Various 2D and 3D change detection techniques have been developed in the literature to monitor changes inside urban areas. Nevertheless, most of theses techniques require the interaction of the user either to input data, set parameters, or to train classifiers. Automatic unsupervised processes have been seldom tackled since they are very difficult to develop if high accuracies are necessary. This paper provides a fully automatic change detection procedure for urban areas monitoring. It exploits at best the information provided by multi-spectral images and Digital Elevation Model (DEM) from two different epochs. A fusion both at feature and decision levels is thus proposed to detect automatically for each epoch the following land cover classes: buildings, shadows, water areas, ground, low and high vegetation. Applying such fusion on Ikonos stereo data acquired over an Asian urban area in spring 2006 and winter 2010 and their ensuing DEMs has proved both the efficiency and worth of the joint use of the multi-spectral and height information. A class-for-class comparison is carried out between the two obtained classification maps to detect the changes that have occurred between 2006 and 2010 over the studied area. A set of standard evaluation measures are finally computed to assess the quality of the proposed procedure.

Keywords: Multi-spectral data, DEM, Information fusion, Automatic 3D change detection.

## Introduction:

In the last few decades, the constantly intensive global urbanization has made the urban and suburban areas among the most dynamic sites on Earth. New innovative tools are thus required for better monitoring of such areas. Remotely sensed imagery in some cases can be a very reliable source for better understanding of urban areas. In fact, Earth observation satellites yielding high resolution data, can provide greatly valuable multi-temporal information that can be further exploited for change detection purposes. Nevertheless, while most of the changes are optically visible and easily detectable by an

expert user, automatic unsupervised processes for change detection are quite difficult to develop. This is mainly due to the fact that urban classes are generally characterized by a very high variability. The semantic class “building” for example presents varying spectral characteristics due to the differences in the specifications of the imaging satellite (e.g. orbit height, orbit inclination, spatial/spectral resolution, etc) and in the materials found on rooftops (e.g. tiles, metal, etc). The monitoring of this main urban class with such varying properties is very difficult if high accuracies are necessary and if the process shall be fully automatic. That is why the interpretation of changes has remained up-to-now visual/manual in most operational applications in remote sensing, requiring the user to stratify manually, set parameters, or to train classifiers.

This paper provides solutions for automatic 3D change detection inside urban areas based on information fusion techniques applied at both feature and decision level. In fact, when urban scenes have to be interpreted, many works have been combining information from different sources to determine plausibility and enforce constraints in label assignments (e.g. (Hoiem et al. 2008), (Chaabouni-Chouayakh and Datcu 2010)). The following land cover categories have been taken into account in our scheme: buildings, shadows, water areas, ground, low and high vegetation. Our solutions attempt to exploit in an intelligent way information extracted from multi-spectral and Digital Elevation Model (DEM) data so that finer and more accurate classification maps can be automatically obtained. Indeed, no spectral information can supply a satisfactory separation between low and high vegetation, or between buildings and ground, due to the similarity of spectral properties between them. For this purpose, this work proposes to fuse the height information extracted from DEM data to the multi-spectral information. In fact, when dealing with urban areas, one possibility to cope with the automatic city growth monitoring is the exploitation of the height information in respect

to the different man-made objects that exist in the scene. The subtraction of DEMs, computed from image pairs acquired at different epochs, should thus provide a valuable information about the 3D urban changes that occurred in the studied area as done in the works of (e.g. (Gong et al. 2000), (Heller et al. 2001) and (Hollands et al. 2007)). However, when at least one of the DEMs exhibits artifacts a simple DEM subtraction will result also in the detection of virtual changes. Thus, the aim of this work is to overcome the stated difficulties and improve unsupervised image classification through an intelligent fusion between the multi-spectral information and DEM data, and apply in a final stage this classification to perform change detection using multi-temporal images. This joint exploitation of height and spatial information has also been the focus of some of our previous publications (e.g. (Chaabouni-Chouayakh et al. 2011) and (Chaabouni-Chouayakh & Reinartz 2011)) and has shown promising change detection results, but the interaction of the user was still required. In this work, a fully automatic unsupervised approach is implemented and evaluated using Ikonos data acquired in spring 2006 and winter 2010 over an Asian urban area.

The organization of this paper is as follows: the different pre-processing steps needed to prepare the data to be used are first described. After that, the recognition of the different urban classes is detailed in the second section. Then, change detection is undertaken and evaluated using different standard measures. Finally, conclusions and some perspectives are provided in the last section.

**Preparation of the data:**

Two pairs of Ikonos-2 stereo images (©EUSI provided under the EC/ESA GSC-DA contract) acquired in spring 2006 and winter 2010 over an Asian urban area have been used. DEMs are generated from each pair of panchromatic stereo images using the *in-*

*house* developed Semi-Global Matching (SGM) algorithm ((Hirschmüller 2008), (d'Angelo et al. 2008)). Then, the resulting DEMs have been interpolated over the areas where the matching between the two stereo images fails (e.g. over occluded areas). Digital Terrain Models (DTMs) are also generated from each DEM using the *in-house* tool of (Arefi et al. 2009). Then, for each epoch the normalized DEM (nDEM) is computed where ground height is referenced to zero providing height information about the above-ground objects independently from the terrain, by subtracting the DTM from the corresponding DEM. Finally, the multi-spectral images are ortho-rectified according to the corresponding DEMs and later pan-sharpened to have the same resolution as the panchromatic images.

**Urban classes recognition:**

The flowchart of the overall classification procedure is given by Fig.1. It is based on a hybrid use of a hierarchical decision-tree and feature extraction/clustering techniques. Several measures and descriptors are computed from both the ortho-rectified pan-sharpened images and the DEM data to identify the different classes.

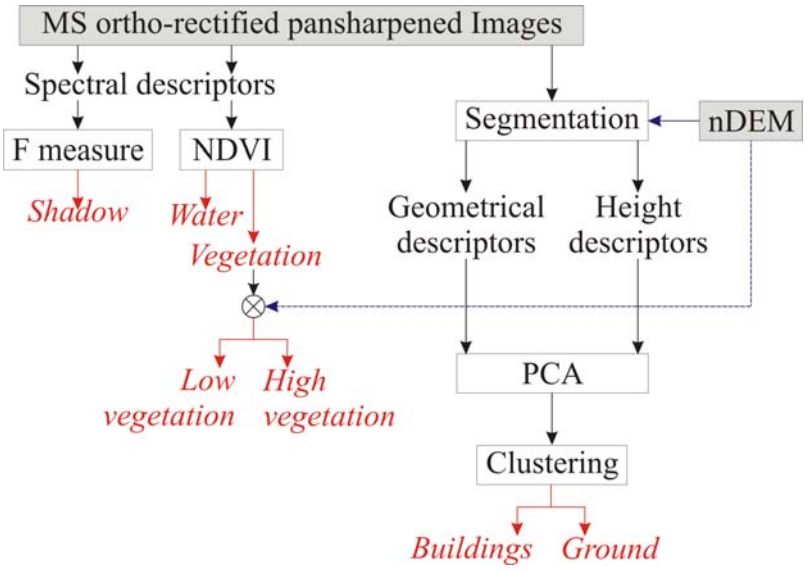
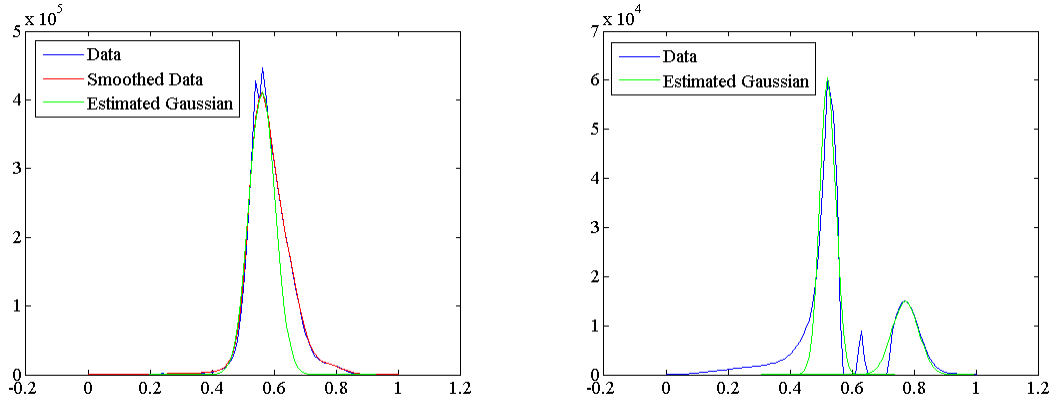



Fig.1: Flowchart of the proposed classification method.

***Vegetation and water recognition***

The Normalized Difference Vegetation Index (NDVI) is calculated as described in (Deering 1978) and automatic histogram thresholding is used to isolate the distributions for vegetation and water bodies, assuming Gaussian distributions. Histogram thresholding of the extracted NDVI values is obtained by average filter smoothing and Gaussian curve fitting. This has been attained in two ways: by localising the first and last Gaussians, corresponding to the targeted areas, and conversely by subtracting the main Gaussian for the enhancement of the rest of distributions. Once a peak is found, a height equivalent to half of the peak is searched for on both sides. The histogram of the NDVI as well as the smoothed histogram and the estimated Gaussian for our study area of Spring 2006 is plotted in Fig.2(a). The localisation of the thresholds used for vegetation and water detection is illustrated in Fig.2(b).




(a) (b)

Fig. 2: Automatic histogram thresholding of the NDVI for vegetation and water detection. (a) Smoothing and Gaussian fitting of the NVDI values. (b) Localising the thresholds for vegetation (here 0.41) and water (here 0.73) detection.

The segments recognized as vegetation are later fused in a decision tree with height information (from DEM data), to distinguish between high vegetation (trees) and low vegetation (grass and bushes).

***Shadow recognition***

The class of shadow is automatically identified by computing the F index formulated by (Marchant & Onyango 2002) and its extension by (Makarau et al. 2011). Again, the histogram of values is plotted, smoothed with a moving average filter, and thresholded by identifying the Gaussian curve corresponding to shadowed areas.

***Buildings and ground recognition***

To detect buildings, nDEM data is not providing a completely reliable information since the data show some artefacts at the borders of the buildings. For better building detection, this work proposes to first apply mean-shift segmentation to the multi-spectral pan-sharpended images using the system developed by (Comanicu & Meer 2002) to group pixels with close spectral properties in the same segment. After that, every segment in the segmented image overlapping with the blobs of the nDEM over a considerable percentage (here set to 60%) is incorporated to the building map. With this, a supervised thresholding of the nDEM is no more needed to detect over-ground objects and a refined nDEM is finally obtained. An example of the improvement achieved after the proposed border refinement is illustrated in Fig.3.

(a) (b) (c)

Fig.3: Refinement of the nDEM: (a) nDEM. (b) Segmented multi-spectral image. (c) Final building map.

After that, several shape features are computed for each segment in the refined building map, as done in the work of (Xie et al., 2011):

- Area represents the size.
- Eccentricity represents the ratio of the distance between the foci of the region best fitting ellipse and its major axis lengths, with values varying from one (maximum elongation of the ellipse) to zero (circular shape). It is defined in (1), where $(\bar{x}, \bar{y})$ is the centroid of the object and $\mu_{pq} = \Sigma\Sigma(x - \bar{x})^p(y - \bar{y})^q$ is the (p,q) order central moment of the shape.

$$E = \frac{\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}} \quad (1)$$

- Solidity stands for the ratio of the area of the polygon and convex hull area of the polygon approximating the shape, with values ranging from one (convex shape) to null.
- Compactness is defined in (2), A and P being respectively the area and perimeter of the region, its values ranging from one (circular shape) to zero (maximally elongated shape).

$$C = \frac{4\pi A}{P^2} \quad (2)$$

- Asymmetry is calculated as in (3) where m and n are respectively the major and minor axis of the ellipse best fitting the area, its values approaching one with higher asymmetry of the region.

$$As = 1 - \frac{n}{m} \quad (3)$$

- Rectangular fit refers to the fit of the region area A to its bounding box area A0, as defined in (4), with values approaching zero for a perfect fit.

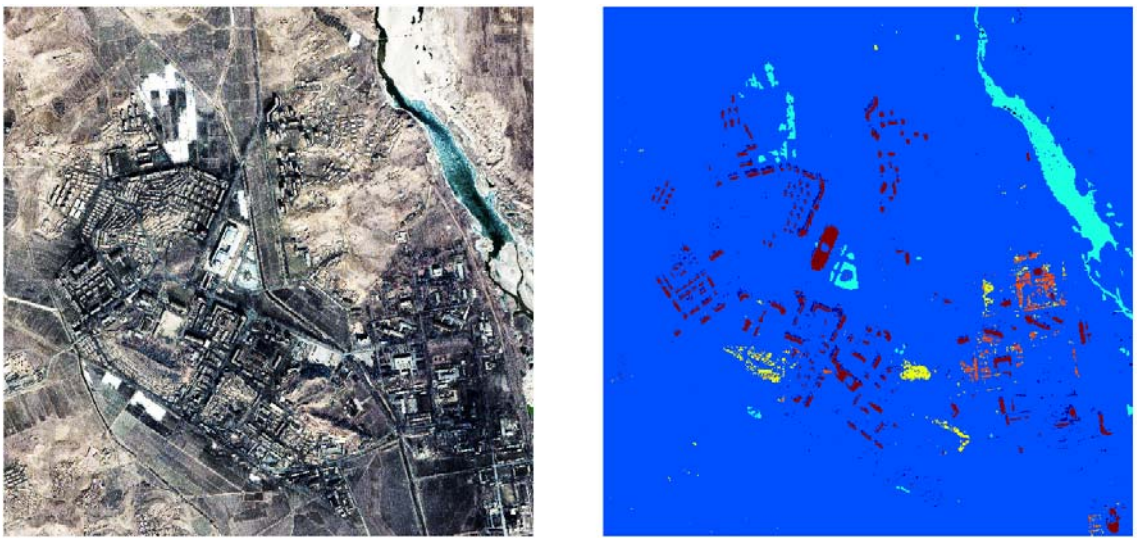
$$Rf = 1 - \frac{A}{A_0} \quad (4)$$

- Length to width ratio is calculated using (5), where a and b are the bounding box length and width, A is the region area, and A0 is its bounding box area.

$$\gamma = \frac{a^2 + ((1 - f)b)^2}{A} \quad \text{where} \quad f = \frac{A}{A_0} \quad (5)$$

Then, a composite is selected with height, and the previously computed shape features to recognize the rest of the classes. Moreover, all Hu moments (Chen et al., 2004) are calculated, but only the first one is considered due to separability reasons. Relevant features are later selected by application of principal components analysis for dimensionality reduction purposes. Finally, K-means clustering is applied on the selected composite to generate classification maps for each epoch.

When applying the previously described procedure on our test data, accurate results are obtained where correspondences between the obtained clusters and desired classes (buildings, shadows, water areas, low vegetation, ground, and high vegetation) are easy to establish. The obtained classification maps for both epochs as well as the corresponding input multi-spectral images are displayed in Fig.4 and Fig.5

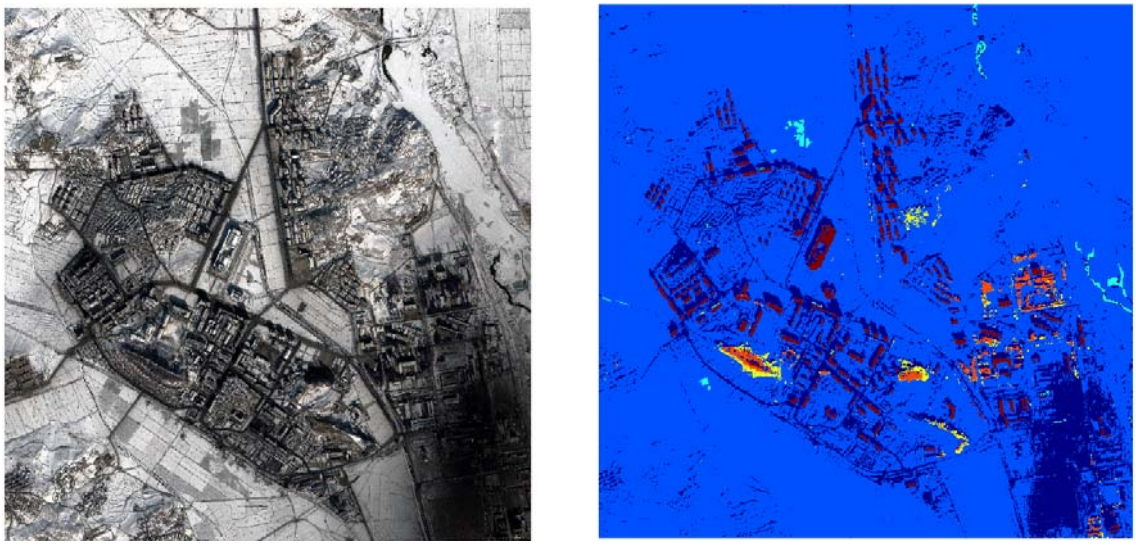


(a) (b)

Fig.4: Classification of the data from Spring 2006: (a) Multi-spectral image. (b) Classification map. The clusters in yellow, orange, red, cyan, blue and dark blue correspond to low vegetation, high vegetation, buildings, water bodies, ground and shadow areas.

**Change detection and evaluation of the proposed method:**

After feature extraction and clustering, class-to-class comparison is carried out between the classification maps obtained at each epoch in order to detect the different changes.



(a) (b)

Fig.5: Classification of the data from Winter 2010: (a) Multi-spectral image. (b) Classification map. The clusters in yellow, orange, red, cyan, blue and dark blue correspond to low vegetation, high vegetation, buildings, water bodies, ground and shadow areas.

In the case of our studied area, the mainly observed 3D changes are linked either to varying levels of vegetation growth since the two stereo images have been acquired in two different seasons (spring and winter), or to building construction since a quite long period (4 years) separates the acquisition years of the two stereo image pairs. The final change map is displayed in Fig.6.

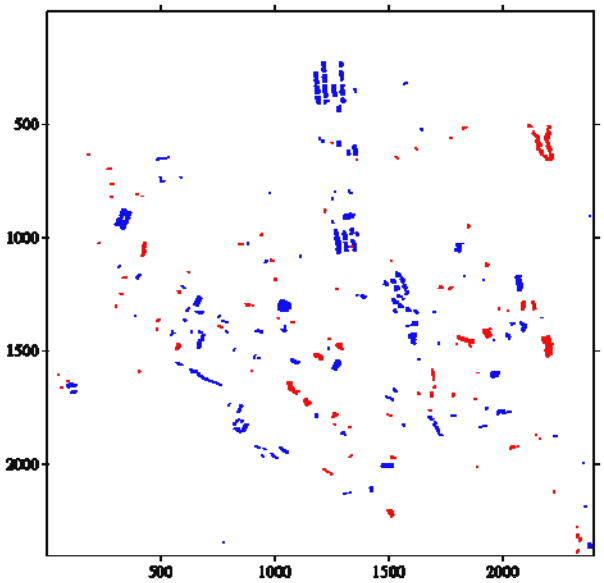



Fig.6: Final change map: The positive and negative changes are highlighted in blue and red respectively.

For quantitative evaluation, a reference comprising an area of 800x800 pixels has been manually created since the whole change map is quite large (2400x2400 pixels). In here, we concentrate on the positive changes since they correspond to new constructions and are therefore easier to sketch. The reference as well as the corresponding change sub-map is displayed in Fig.7. A visual comparison shows already that most of the changes are successfully retrieved. Some virtual changes have been wrongly detected but they correspond to very small objects in comparison to the real changes.

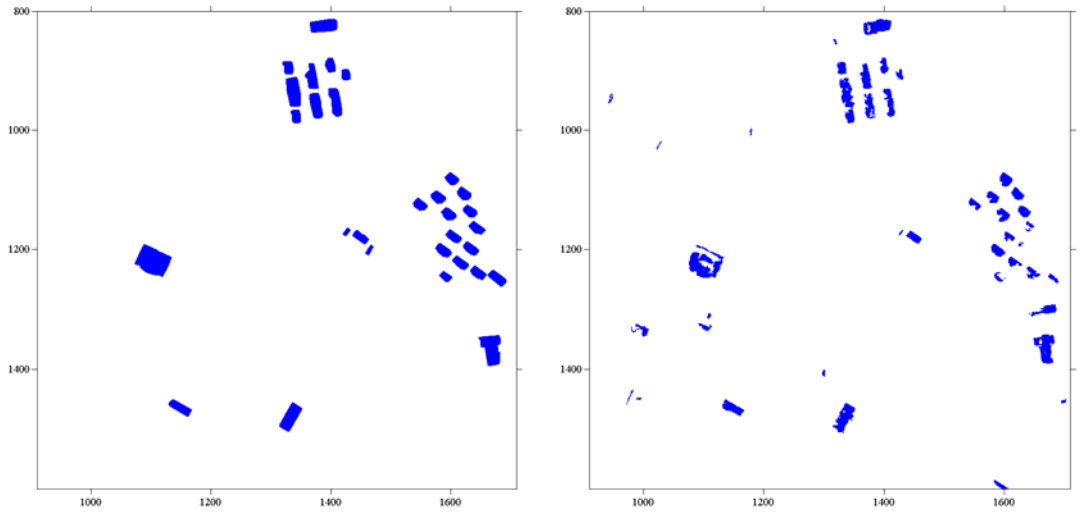



Fig.7: Reference (left) versus change detection results (right) comparison.

Finally, the following objective metrics widely used in the literature (e.g. (Sohn & Dowman 2007) and (Ekhtari et al. 2008)), were computed:

$$\begin{aligned} \text{Branching Factor} &= \text{FP}/\text{TP}, \\ \text{Miss Factor} &= \text{FN}/\text{TP}, \\ \text{Completeness}(\%) &= 100 \times \text{TP}/(\text{TP}+\text{FN}), \\ \text{Correctness}(\%) &= 100 \times \text{TP}/(\text{TP}+\text{FP}), \\ \text{Quality Percentage}(\%) &= 100 \times \text{TP}/(\text{TP}+\text{FN}+\text{FP}) \end{aligned} \tag{6}$$

The variables TP, FP, TN and FN refer respectively to true positives, false positives, true negatives and false negatives for a two-class comparison (here “change” and “non-change”). The evaluation measures have been computed at pixel and object levels. The results are summarized in Tab.1. We can notice that our algorithm performs differently

at pixel and object levels. In fact, at object level, it tends to produce less FN objects than FP ones, resulting in a miss factor of 0.03 and branching factor of slightly poor performance (0.43). Also, the completeness rate is higher than the correctness one. This indicates that the number of missed changes is less than the number of over-classified changes. However, at pixel level, the algorithm produces less FP pixels than FN ones which results in a correctness rate (almost 86%) higher than the completeness one. Finally, the proposed change detection technique could reach a quality percentage of more than 60 and 68% at pixel and object levels, respectively, proving its efficiency and relatively high accuracy, especially that it is fully automatic.

Tab.1: Quantitative evaluation of the proposed change detection procedure.

| | Branching factor | Miss factor | Completeness | Correctness | Quality percentage |
|---|---|---|---|---|---|
| Pixel level | 0.16 | 0.49 | 67.09% | 85.79% | 60.39% |
| Object level | 0.43 | 0.03 | 96.77 % | 69.76 % | 68.18 % |

**Conclusions & perspectives**

This paper provides a fully automatic 3D change detection procedure through a joint use of spectral and height information. It is a two-step procedure. Firstly, several measures and descriptors are computed from both multi-spectral images and DEM data to recognize the following urban classes: buildings, shadows, water areas, ground, low and high vegetation. Secondly, class-for-class comparison between the classification maps obtained from two epochs is carried out to detect the different changes. The approach is tested and evaluated using Ikonos data. A number of common object-based metrics (branching factor, miss factor, completeness, correctness, quality percentage) are computed showing the effectiveness of the developed algorithm.

The results show a real progress toward automatic 3D change detection inside urban

areas, although some of them might still be improved and completed. They can be considered as preliminary results for some higher level applications, such as refining urban area monitoring through a more intensive exploitation of the multi-spectral information so that a wider range of man-made structures is involved. Also, DEMs could be further processed to identify several levels of building or vegetation growth.

## Bibliography


Arefi, H., d'Angelo, P., Mayer, H. and Reinartz, P., 2009. Automatic Generation of Digital Terrain Models from Cartosat-1 Stereo Images. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.

Chaabouni-Chouayakh, H. and Datcu, M., 2010. Coarse-to- Fine Approach for Urban Area Interpretation Using TerraSAR-X Data. IEEE GRSL 7(1), pp. 78–82.

Chaabouni-Chouayakh, H., d'Angelo, P., Kauss, T. & Reinartz, P., 2011: Automatic Urban Area Monitoring Using Digital Surface Models and Shape Features, Joint Urban Remote Sensing Event, Munich, Germany.

Chaabouni-Chouayakh, H. and Reinartz, P., Towards Automatic 3D Change Detection inside Urban Areas by Fusing Height and Shape Features, PFG, 2011.

Q. Chen, E. Petriu, and X. Yang. A Comparative Study of Fourier Descriptors and Hu's Seven Moment Invariants for Image Recognition. In Canadian Conference on Electrical and Computer Engineering 2004 (IEEE Cat. No.04CH37513), pp. 103-106, Niagara Falls, Ont., Canada, 2004.

Comanicu, D. and Meer, P. 2002: Mean Shift: a Robust Approach Toward Feature Space Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24:603-619.

d'Angelo, P., Lehner, M., Krauss, T., Hoja, D. & Reinartz, P., 2008: Towards Automated DEM Generation from High Resolution Stereo Satellite Images", Proceedings of ISPRS, Peking, China, pp. 1137-1342.

Deering, D.W.. 1978, Rangeland Reflectance Characteristics Measured by Aircraft and Spacecraft Sensors. PhD thesis, Texas A&M University.

Ekhtari, N., Sahebi, M.R., Valadan Zoej, M.J. & Mohammadzadeh, A., 2008: Automatic Building Detection From Lidar Point Data", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37 (B4), Beijing, China.

Gong, P., Biging, G.S. & Standiford, R., 2000: Use of Digital Surface Model for Hardwood Rangeland Monitoring, Journal of Range Management, **53** (6**)**: 622-626.

Heller, A. J., Leclerc, Y. G. & Luong, Q. T., 2001: A Framework for Robust 3-D Change Detection, Proceedings of SPIE, Toulouse, France.

Hirschmüller, H., 2008: Stereo Processing by Semiglobal Matching and Mutual Information, IEEE Transactions on Pattern Analysis and Machine Intelligence, **30** (2): 328-341.

Hoiem, D., Efros, A. A. and Hebert, M., 2008. Closing the Loop on Scene Interpretation . CVPR.

Hollands, T., Boström, G., Goncalves, J.G.M., Gutjahr, K., Niemeyer, I. & Sequeira, V., 2007: 3D Scene Change Detection from Satellite Imagery, Proceedings of the 29th Symposium on Safeguards and Nuclear Material Management, Aix-en-Provence, France, pp. 1-6.

Makarau, A., Richter, R., Muller, R., and Reinartz, P. 2011: Adaptive Shadow Detection using a Blackbody Radiator Model. IEEE Transactions on Geoscience and Remote Sensing **49**(6): 2049-2059.

Marchant, J. A. and Onyango, C. M.. 2002: Spectral Invariance under Daylight Illumination Changes. Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 19(5):840-848.

Sohn, G. & Dowman, I., 2007: Data Fusion of High-resolution Satellite Imagery and LiDAR Data for Automatic Building Extraction", ISPRS Journal of Photogrammetry and Remote Sensing 62 (1): 43-63.

Xie, H., Heipke, C., Lohmann ,P., Soergel, U., Tong, X. and Shi, W., 2011: Ein neuer Binarcodierungs-Algorithmus zur gemeinsamen regionenbasierten Klassifikation von Hyperspektraldaten und digitalen Oberachenmodellen. Photogrammetrie – Fernerkundung – Geoinformation, pp. 17-33.