BUILDING DETECTION WITH MULTI-VIEW COLOUR INFRARED IMAGERY

Charles Beumier, and Mahamadou Idrissa

Royal Military Academy, CISS Department, Brussels, Belgium;
{beumier / idrissa}(at)elec.rma.ac.be

ABSTRACT

This paper presents an automatic building detection approach exploiting colour infrared (CIR) imagery. A Digital Surface Model (DSM) is first extracted by photogrammetry with a proprietary development including multiple views. Then a Digital Terrain Model (DTM) is derived from the DSM by selecting the lowest regions with slowly varying elevation. A normalised DSM, difference of the DSM and derived DTM highlights building candidates thanks to a simple threshold. In order to reduce false positives due to trees, a vegetation mask is obtained from the *NDVI* of the ortho-rectified CIR image. A specific procedure was designed to handle the problematic shadow areas where the NDVI criterion often misses vegetation.

The proposed building detection approach has been applied to the Vaihingen data set of the ISPRS benchmark and validated by the results of its evaluation procedure. We plan to integrate the new DTM extraction and shadow adapted vegetation mask, main contributions of this paper, to our project about change detection for database revision for the Belgian National Geographic Institute.

INTRODUCTION

In order to deliver precise cartographic data, National Mapping Agencies (NMAs) need to find changes in the land cover by identifying built zones, road elements, vegetation areas, etc. The traditional update by human interpreters becomes too expensive in time and money so that NMAs are now trying to integrate semi-automatic procedures in their production line. This is exemplified by projects like ATOMI in Switzerland (1), DB TOP10DK in Denmark (2), MAJEC in France (3) and WIPKA in Germany (4).

As presented by Baltsavias (5), a large variety of computer assisted building detection approaches have been designed, depending on the source data (aerial/satellite images, panchromatic/multispectral data, photogrammetric/LIDAR surface models) and the type of features exploited (radiometric, geometrical, spectral, contextual).

Roughly speaking, early developments mostly considered land cover classification or simple geometrical features (line, corner or rectangle in (6)). Later studies integrated several types of features to improve the robustness of building detection. For example Müller (7) uses spectral data, the shadow proximity, many geometrical features and contextual classification by incorporating surrounding building hypotheses.

Nowadays most developments exploit the third dimension, as exemplified in the approaches adopted for the EuroSDR test about automatic building change detection (8). In the theses of Niederöst (1) and Champion (3), the elevation feature of a DSM is normalised by a DTM to obtain the so-called normalized DSM ("nDSM") representing local elevation. The thresholded nDSM offers building candidates which are filtered out by a vegetation mask obtained from spectral data to reject false candidates corresponding to trees.

The work presented here focuses on building detection and responds to one of the themes of the ISPRS benchmark on urban object detection and 3D building reconstruction (9). As depicted in Figure 1, we followed an approach similar to the one summarized in the previous paragraph, although we have not used any building reference. We exploited the original aerial images in order to

extract a DSM with a proprietary method and derive an ortho-rectified colour infrared (CIR) image. An original DTM from DSM extraction procedure has been developed to derive a nDSM used to highlight building candidates. The rejection of vegetation was made pixel-wise thanks to the *NDVI* vegetation index with a special treatment for shadow areas.
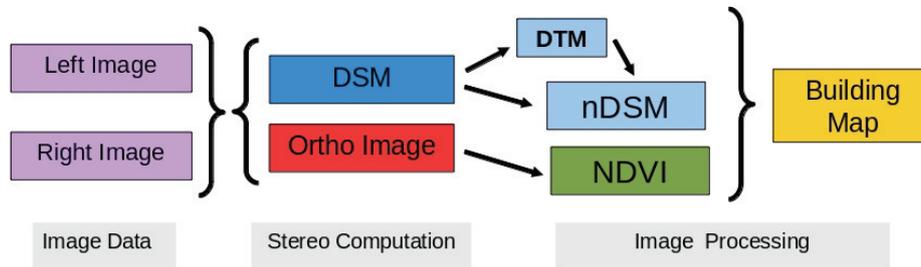

Figure 1: Approach overview for building map creation.

27 participants sent their results to the ISPRS benchmark organizers for evaluation. As summarised in the paper presenting the results (9), the processing strategies may be classified into three groups: model-based, not using training (18 methods); supervised classification based on training data (7 methods); and statistical sampling using heuristic models (2 methods). Like three other participants, our solution is model-based and only relies on original images (stereo pair), not using the provided LIDAR DSM, image DSM and ortho image (see next section). The difference between the four model-based approaches with original images lies in the way the DSM and DTM were produced and the way the nDSM and vegetation index were combined to create the building map. One method used Dempster-Schafer, another eCognition and the third proceeded by vegetation elimination as we did. The interested reader is invited to consult references given in (9).

METHODS

Reference data

The data used for development and test is the set of Vaihingen of the ISPRS benchmark (10). The imagery consists of a set of CIR aerial images, a derived DSM map and an ortho-rectified CIR image with 9 cm spatial resolution. A DSM has also been produced by laser scanning at 25 cm resolution but we have not considered this source since we wanted to validate our complete processing chain as depicted in Figure 1, exploiting the original CIR aerial images only.

DSM

In order to produce a DSM, we followed the classical approach of stereo matching (11) consisting of epipolar transformation, disparity estimation and 3D reconstruction. Our development took advantage of the high radiometric resolution (11-bit) and the multiplicity of image pairs (multi-view).

Epipolar rectification is the transformation of a stereoscopic pair into two images so that the two image pixels corresponding to the same scene point have an identical coordinate. A set of SIFT points (12) were automatically detected in each original image of the pair and associated by a RANSAC procedure (13) to find the epipolar transformation using the 8-point algorithm (14). The epipolar images simplify the search for matching pixels since they have an identical coordinate. Matching pixels becomes an efficient 1-D search, leading to the notion of disparity, equivalent to the distance of the matching pixels.

In order to extract a dense DSM, we need to evaluate the disparity at each pixel, if possible. For better robustness, pixel matching was performed by correlation of an area surrounding the pixel. The best area size depends on the image content. Instead of considering size adaptation, we preferred to compute several disparity maps for a few different area sizes, from 3×3 to 200×200 pixel. Each pixel had thus a small set of disparity candidate values. To select the best disparity of a pixel, we collected the disparity sets of all pixels in a neighbourhood limited in size and by an edgemap. The neighbourhood offers a better statistical estimation while the edgemap limits the influence of

disparity values from other objects. The median value of this collection was the disparity value chosen for the pixel. This approach provides much better results than a fixed-size window technique and was developed since global approaches are computationally demanding.

The DSM is obtained by 3D reconstruction, using the collinearity equations to compute geographic coordinates (*X*,*Y*,*Z*) of any point from its image coordinates (*x*,*y*) and disparity value. This image to world coordinate transformation requires the camera intrinsic parameters (focal length, optical centre and sensor size) and the camera position and orientation for each single image. In multi-view stereo, surface points may be covered by several stereo pairs. The DSMs of all these pairs were computed and combined into a single DSM using the median function that is particularly robust to outlier values.

The developed DSM extraction has been tested in the context of the benchmark about dense image matching organised by EuroSDR. The DSMs produced by a tenth of participants for a common dataset with commercial or proprietary solutions are compared in the paper of Haala (15). Our development (called RMA) appears as a solution with average quality, although no ranking of the methods was performed by the benchmark.

Figure 2a shows a part of the DSM computed for the Vaihingen/Enz data set. As can be seen, the use of the proposed algorithm provides geometrically accurate DSM values for spatially defined objects (building boundaries) as well as textureless regions. The availability of multiple views makes the DSM nearly free from holes (where there is no image match in any stereoscopic views).
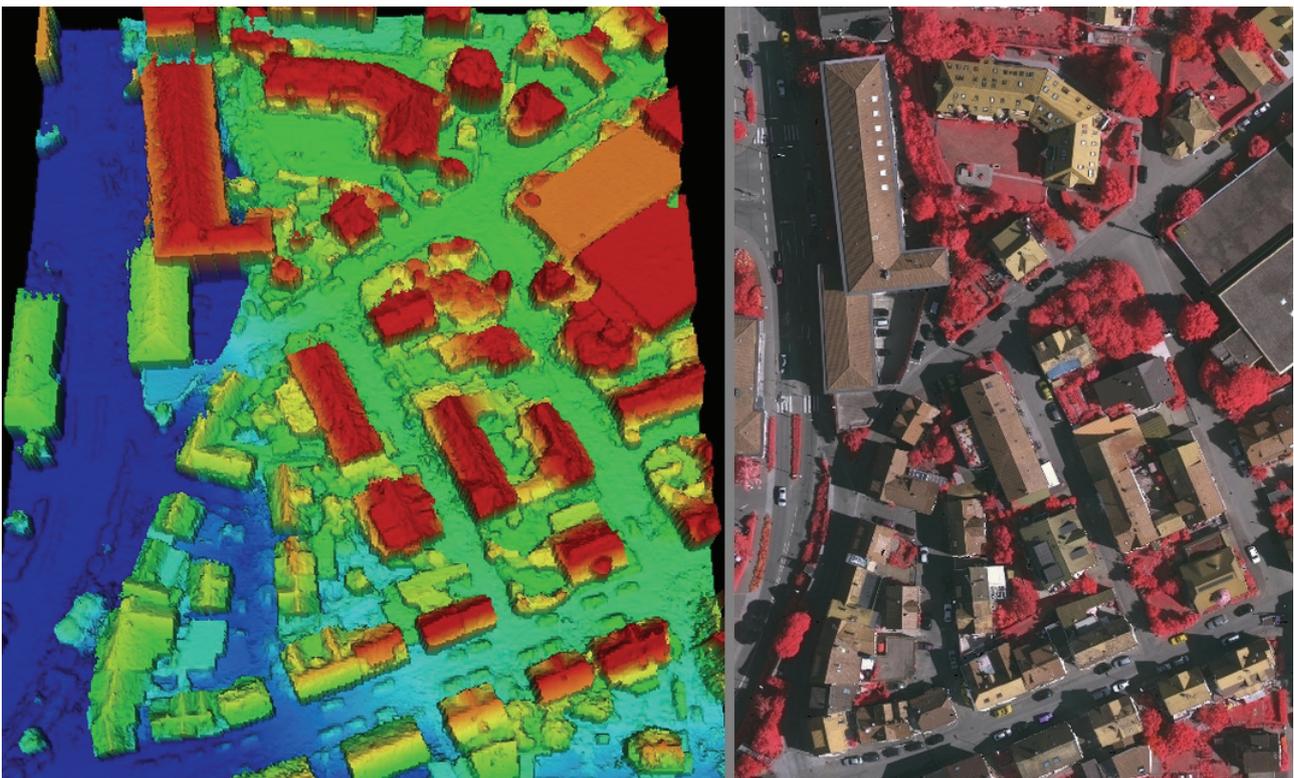

Figure 2: (a) Oblique view of the DSM; (b) CIR ortho-rectified image.

Ortho-rectification

Ortho-rectification is the geometrical correction that allows true distances to be measured on the image. Thanks to sensor parameters and the use of collinearity equations, object space coordinates (*X*,*Y*) corresponding to each pixel in the final ortho-image and the corresponding elevation *Z* (from the DSM) are projected into the source image to obtain the CIR value for the ortho-image pixel at (*X*,*Y*). This surface to image projection produces some artefacts in the ortho-image like double mapping due to occluded areas. The second mapping is avoided using each original image pixel once for the corresponding highest *Z*-value. This is implemented thanks to a *Z*-buffer. The

remaining holes in image (no value due to occlusion) are limited thanks to the multiple-view imagery. Figure 2b shows an ortho-image for Area1 of the ISPRS test data.

DTM from DSM

A common approach for building detection consists in thresholding the normalized DSM obtained by subtracting a DTM to the DSM (nDSM=DSM-DTM), and corresponding to the local elevation of non-ground objects like houses, trees or bridges. In case no DTM is available, the best solution consists in deriving the DTM from the DSM, as suggested by Champion (3).

Our DTM from DSM approach consists of three steps: DSM segmentation into uniform regions, rejection of regions higher than neighbours and interpolation in rejected regions. Details and results for the same dataset are given in (16).

Vegetation Mask

The elevation cue is a strong indicator for building presence but it suffers from false positives corresponding to elevated vegetation (trees). The traditional solution consists in applying a vegetation mask usually obtained by a threshold on the Normalized Difference Vegetation Index computed at each pixel by $NDVI = (IR-R)/(IR+R)$, where IR and R represent respectively the infrared and red intensities.

Even if the $NDVI$ is a strong cue it suffers from the poor spectral resolution in the shadow areas where most objects appear greyish so that the $NDVI$ tends to 0, possibly leading to wrong building hypothesis. We observed that the presence of vegetation was perceptible when the intensity of the IR channel was sufficiently higher than the Red and Green intensities.
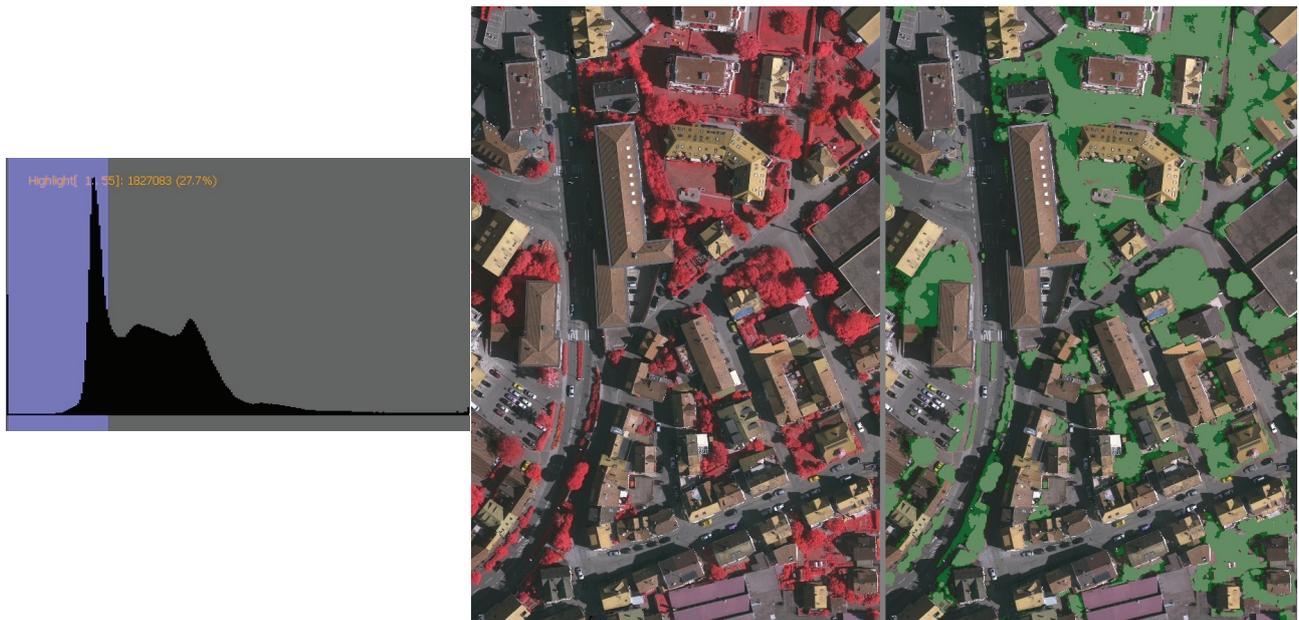

Figure 3: (a) Shadow threshold from red histogram; (b) CIR ortho image; (c) Vegetation mask (in green: NDVI > 0.2) enlarged by vegetation in shadow (in dark green: new rule).

We adapted the vegetation mask to include those shadow pixels with IR intensity at least 10 percent higher than the mean of the Red and Green intensities. A pixel is declared in shadow if its Red intensity belongs to the first peak of the Red histogram. The threshold for shadow is chosen as the first maximal negative slope (on the right of the first peak, Figure 3a).

We see in Figure 3c a part of Area1, with vegetation detected by $NDVI > 0.2$ (in green) and vegetation in shadow detected by the new rule ($IR >= 1.10 \cdot (R+G)/2$, in dark green). We notice that the new rule partially completes vegetation detection in shadow areas. This is particularly useful for trees in shadow that would lead to false positives if they were not recognized as vegetation.

Building Map

The building map is made of nDSM pixels with sufficient height, but out of the vegetation map. Holes in the building map (for instance due to reflections by roof windows) are handled by a morphological closure with a square structural element of size equal to 1m. Spurious protrusions are eliminated by a morphological opening with a square structural element of size equal to 2 m.

RESULTS

The proposed building detection approach has been evaluated independently according to the procedure described in (17) for the Vaihingen dataset of the ISPRS benchmark. The building map for Area1, Area2 and Area3 were sent for evaluation. We received back a complete analysis described in (18) with figures and images. Building detection can be qualitatively evaluated from Figure 4, where yellow regions are correctly identified as buildings, red regions stand for false positives (wrong buildings) and blue regions correspond to false negatives (missing buildings).
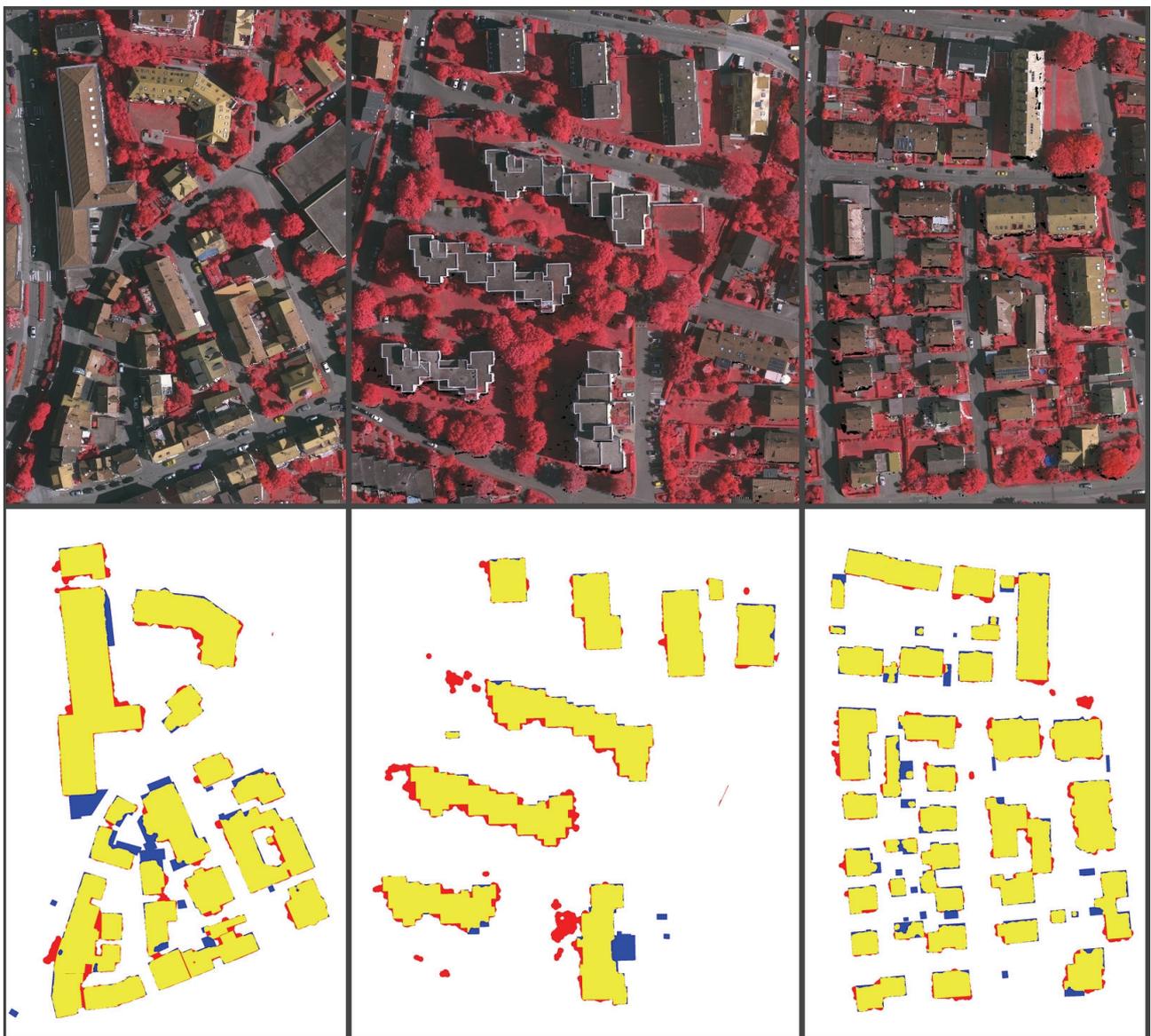

Figure 4: Evaluation of building detection. (a) Top row: CIR ortho-rectified image parts for Area1, Area2 and Area3. (b) Bottom row: Yellow = correct identification, Red = wrong detection, Blue = missing building parts.

Table 1 presents quantitative building detection results in terms of completeness and correctness for the three areas, per object and per area (pixel). In per-object evaluation, an object (connected set of pixels) is considered detected if at least half of its area is covered by building pixels in the reference. Completeness is the ratio of the reference objects (or pixels) detected as building. Missing buildings (False Negatives) degrade completeness. Correctness is the ratio of correct detection of building objects (or pixels) relatively to all detected objects (or pixels). The False Positives (detection which are not real buildings) degrade correctness. The quality factor is a ratio comparing reference objects (or pixels) to both kinds of errors (false positives and false negatives).

With *TP*, *FP* and *FN* respectively standing for the number of true positive, false positive and false negative items, we have:

$$Completeness = TP / (TP + FN)$$

$$Correctness = TP / (TP + FP)$$

$$Quality = TP / (TP + FP + FN)$$

Table 1: Quantitative results of building detection per pixel and per object for Area1, Area2 and Area3.

Evaluation method	Area1	Area2	Area3
Per-area Completeness	91.6%	95.4%	91.3%
Per-area Correctness	92.4%	85.9%	92.4%
Per-area Quality	85.2%	82.6%	84.9%
Per-object Completeness	83.8%	85.7%	78.6%
Per-object Correctness	96.9%	52.2%	93.9%
Per-object Quality	81.6%	48.0%	74.7%

The evaluation on Area1 mainly reveals missing building parts (in blue in Figure 4). The largest missing part concerns a building with flat roof between two roads of different elevation. The flat roof at the level of the right road was kept as ground area and is part of the DTM. The other missing parts are mainly due to shadow problems or non optimal threshold (2 m). The largest false positive (in red) concerns a shadow area between a house and trees at the bottom left of the image. The DSM obtained by image matching was wrongly estimated in this area (the LASER scanning DSM seems better in that region).

In Area2, one main building part is missing, due to vegetation on a flat roof. Other missing building areas are small. A lot of shadow regions were detected as building candidates (in red). As some of these wrong building detections concern isolated regions, they are counted as False Positive objects. Relatively to the low number of objects in this area, they imply poor per-object correctness (52.2%), although their total impact on the per-area correctness (85.9%) is limited.

In Area3, many small buildings were missed (in blue). They correspond to low structures like garages and garden huts. A lower height threshold would help detecting them, but we kept the same threshold (2.0 m) as a good compromise for the three areas. The main false candidate (top right) is due to wrong DSM estimation in a dark area close to trees. Other false positive areas are quite small and relate to shadow regions.

DISCUSSION

The quantitative results of our approach compare to other participant results of the ISPRS benchmark (9). This confirms the correctness of the proposed developments concerning the DSM estimation, the DTM from DSM extraction and the adaptation of the vegetation mask in shadow areas. The reader interested in the benchmark and its results with a categorization of participant

approaches is invited to consult the paper of Rottensteiner (9).

It is worth noting that unlike some other methods, the proposed approach is unsupervised, meaning that no reference data was used for learning. The parameters were chosen following common sense and results were visually compared to the image for rough validation. In particular, the height threshold on the nDSM was set to 2.0 m for all three areas to target building detection down to garages and garden huts. We tried other values and realized that 2.5 m better served Area2 polluted by false positives. Area3 is better handled with 1.7m to detect more low buildings (in blue in Figure 3). The influence of the *NDVI* parameter is limited as long as it remains in the range [0.1,0.3].

With an approach solely based on image (DSM, DTM and vegetation mask), the most difficult parts for building detection concerns shadow areas where image data has poor spectral definition. Vegetation on roof appears to be a new challenge.

From the different image processing operations, the DSM extraction is the most computationally expensive, especially since whole (huge) images were processed. Once the DSM and the ortho image are available, the building map is produced in about 10 seconds (PC with Intel Core 2 Duo @ 2.33 GHz, 2 GB RAM) for the images of the test areas (about 4Mpix each).

CONCLUSIONS

We have presented an approach for building detection from CIR multi-view imagery. It follows the classical local elevation and vegetation mask principle. First a DSM is obtained from a proprietary development exploiting multi-view image matching. This DSM is normalized by a DTM directly derived from the DSM by an original method selecting lower areas with slowly varying elevation and interpolating between these areas. Secondly, a vegetation mask is constructed by keeping pixels with high *NDVI*. This mask is adapted to include vegetation in shadow areas where the *NDVI* is usually underestimated. The building mask is the thresholded nDSM filtered by the vegetation mask to exclude elevated vegetation like trees.

The method is quite simple and very fast if the time needed for DSM computation is excluded. It seems to capture a large deal of the image potential when quality scores are compared to those of other methods of the ISPRS test. Most errors arise in the shadow areas where image matching (DSM) and spectral values (vegetation mask) suffer from the limited radiometric resolution.

Many errors are located along the building borders, especially on shadow sides. Image geometrical features like edges, not yet included in the approach, would help refining building borders.

ACKNOWLEDGEMENTS

REFERENCES

1   Niederöst M, 2003. Detection and Reconstruction of Buildings for Automated Map Updating. PhD Thesis. Mitteilungen / Institut für Geodäsie und Photogrammetrie an der Eidgenössischen Technischen Hochschule Zürich, Nr. 78, 155 pp.

2   Olsen B, 2004. Automatic Change Detection for Validation of Digital Map Databases. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2004, XXXV Part B2: 569-574

3   Champion N, 2011. Détection de changement 2D à partir d'imagerie satellitaire. Application à la mise à jour des bases de données géographiques. PhD Thesis, Université Paris Descartes, 178 pp.

4   Busch A, M Gerke, D Grünreich, C Heipke, C-E Liedtke & S Müller, 2004. Automated verification of a topographic reference dataset: System design and practical results. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXV Part B2: 735-740

5   Baltsavias E, 2004. Object extraction and revision by image analysis using existing geodata and knowledge: Current status and steps towards operational systems. International Journal of Photogrammetry and Remote Sensing, 58: 129-151

6   Shufelt J, 1999. Performance evaluation and analysis of monocular building extraction from aerial imagery. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(4): 311-326

7   Müller S & D Zaum, 2005. Robust building detection in aerial images. Proceedings of the ISPRS Workshop CMRT 2005, XXXVI-3/W24, 143-148

8   Champion N, F Rottensteiner, L Matikainen, X Liang, J Hyyppä & B Olsen, 2009. A test of automatic building change detection approaches. Proceedings of the ISPRS Workshop CMRT09, XXXVIII-3/W4, 145-150

9   Rottensteiner F, G Sohn, M Gerke, JD Wegner, U Breitkopf & J Jung, 2014. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing, 93: 256-271

10   Rottensteiner F, G Sohn, M Gerke & J-D Wegner, 2013. ISPRS test project on urban classification and 3D building reconstruction (ISPRS Commission III – Photogrammetric Computer Cision and Image Analysis) 17 pp. (last date accessed: 30 Oct 2014)

11   Scharstein D & R Szeliski, 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision, 47(1-3): 7-42

12   Lowe D G, 1999. Object recognition from local scale-invariant features. Proceedings of the Seventh IEEE International Conference on Computer Vision, 2: 1110-1157

13   Fischler M & R Bolles, 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the Association for Computing Machinery, 24: 381-395

14   Hartley R & A Zisserman, 2004. Multiple View Geometry in Computer Vision, 2nd edition (Cambridge University Press) 670 pp.

15   Haala N, 2013. The landscape of dense image matching algorithms. Photogrammetric Week '13, edited by D Fritsch (Wichmann, Berlin/Offenbach) 271-284

16   Beumier C & M Idrissa, 2014. Deriving a DTM from a DSM by uniform regions and context. Submitted to EARSeL eProceedings

17   Rutzinger M, F Rottensteiner & N Pfeifer, 2009. A comparison of evaluation techniques for building extraction from airborne laser scanning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2(1): 11-20

18   Rottensteiner F, 2013. ISPRS Test Project on Urban Classification and 3D Building Reconstruction: Evaluation of Object Detection Results (ISPRS WG III/4) (last date accessed: 21 Nov 2014)

19   Cramer M, 2010. The DGPF test on digital aerial camera evaluation – overview and test design. Photogrammetrie, Fernerkundung, Geoinformation, 2: 73-82