A TEST OF 2D BUILDING CHANGE DETECTION METHODS: COMPARISON, EVALUATION AND PERSPECTIVES

Nicolas Champion[a], Leena Matikainen[b], Franz Rottensteiner[c], Xinlian Liang[b], Juha Hyyppä[b]

[a]IGN/MATIS, Saint-Mandé, France - nicolas.champion@ign.fr
[b]FGI, Dept. of Remote Sensing and Photogrammetry, Masala, Finland – (leena.matikainen, xinlian.liang, juha.hyyppa)@fgi.fi
[c]Cooperative Research Centre for Spatial Information, Dept. of Geomatics,University of Melbourne, Australia - franzr@unimelb.edu.au

Commission VI, WG VI/4

KEY WORDS: Change Detection, 2D Vector Databases, Algorithms Comparison, Quality Assessment

ABSTRACT:

In the past few years, 2D topographic databases have been completed in most industrialised countries. Most efforts in National Mapping and Cadastral Agencies (NMCAs) are now devoted to the update of such databases. Because it is generally carried out manually, by visual inspection of orthophotos, the updating process is time-consuming and expensive. The development of semi-automatic systems is thus of high interest for NMCAs. The obvious lack of expertise in the domain has driven EuroSDR to set up a test comparing different change detection approaches. In this paper, we limit the scope of the project to the imagery context. After describing input data, we shortly introduce the approaches of the working groups that have already submitted results. Preliminary results are assessed and a discussion enables to bring out first conclusions and directions.

1. INTRODUCTION

2D topographic databases have been completed in most National Mapping and Cadastral Agencies (NMCAs) during the last decade. The main issue now concerns the map revision. This procedure is known to be very tedious, time-consuming and expensive. There is also a growing need to automate it. The development of semi-automatic tools that are able to detect the changes in a database from recent data (typically imagery or LIDAR) and to present them to a human operator for verification is therefore highly desirable. Only a few solutions have been proposed by academic research, even fewer by private companies. Many questions that have arisen remain unanswered, e.g. those regarding the most efficient methodology, the type of primary data to use (LIDAR / imagery) or the most appropriate spatial resolution to choose. These considerations have driven EuroSDR (European Spatial Data Research - http://www.eurosdr.net) to set up a change detection project. This project is in line with previous EuroSDR projects, e.g. the project on road updating (Mayer et al., 2006) and the one on change detection (Steinnocher and Kressler, 2006). The aim of this new project is to evaluate the feasibility of semi-automatically detecting changes in a 2D building vector database from imagery or LIDAR. Three specific topics are investigated in detail: firstly, the impact of the type of data and methodology on the performance of the change detection; secondly, the impact of the spatial resolution of input data; finally, the impact of the complexity of the scene, especially with respect to topography and land use. The methodology of the project consists in a test comparing five different algorithms that are representative of the current state-of-the- art in the field of change detection. It is the main goal of this EuroSDR project to gather enough experience to identify key problems in change detection and to give promising directions for building an optimal operational system in the future.

In this paper, preliminary results achieved for three different algorithms, (Matikainen et al., 2007), (Rottensteiner, 2007) and (Champion, 2007), are presented. In Section 2, the datasets and the comparison method are described. In Section 3, the three approaches of the working groups that have already submitted results are shortly introduced. Results are given and evaluated in Section 4. We finally present a summary and conclusions.

2. INPUT DATA AND TEST SET-UP

2.1 Datasets Description

Two test areas are used in this study. The first test area is situated in Marseille (France). It has an area of about 0.4 km^2 and contains about 1300 buildings. The area corresponds to a very dense urban settlement and features a complicated urban configuration (lower buildings connected to higher buildings). The test area is hilly (with height differences of 150 m) and vegetated, especially along streets. The second test area is situated in Toulouse (France). It has an area of about 1 km^2 and contains about 200 buildings. It features a suburban area and is composed of detached buildings that are very different to each other with respect to the size, height, shape and roofing material. The terrain is also undulating (with height differences of 100 m) and vegetated. In this study, colour infrared (CIR) aerial images with a Ground Sample Distance (GSD) of 20 cm and multiple overlap (a forward and a side lap of minimum 60%) are used for Marseille. Pléiades tri-stereoscopic Satellite CIR images are used for Toulouse, with a GSD equal to 50 cm. In both cases, a Digital Surface Model (DSM) was computed using a stereo-matching algorithm based on the 2D minimization of a cost that takes into account discontinuities and radiometric similarities (Pierrot-Deseilligny and Paparoditis, 2006). The GSD of the DSM is equal to the GSD of the aerial images. CIR orthophotos were also computed from input DSM and images. Reference (up-to-date) building databases were edited manually, by field surveying. Out-of-date databases were derived by simulating changes to the reference databases by inserting new and deleting some of the existing buildings. In Marseille, 107 changes were simulated (89 new and 18 demolished buildings) and 40 changes (23 new and 17 demolished buildings) were simulated in Toulouse. The out-of-date databases were converted to binary image files (building vs. no building) having the same GSD as the input data. These binary building masks were distributed to the participants along with the CIR orthophotos and the DSMs.

2.2 Evaluation Procedure

Each group participating in the test was asked to deliver a change map in which each building is labelled either as *unchanged*, *demolished* or *new*. However, both the representation of the results of change detection and the output formats varied considerably between the individual algorithms. In addition, the definitions of the classes that are discerned in the change detection algorithms are not identical. Whereas the algorithm by Champion (2007) exactly matches the test requirements, this is not the case for the other two algorithms used in this study. In these two cases it was thus decided to use a building label image representing the updated building map for the evaluation. The evaluation consists of a comparison of the outcomes of each algorithm to ground truth (i.e. the initial reference database). Two quality measures are computed for the evaluation: the *completeness*, i.e. the percentage of the actual changes that are detected by an algorithm, and the *correctness*, i.e. the percentage of the changes detected by an algorithm that correspond to real changes (Heipke et al., 1997):

$$\begin{aligned} Completeness &= \frac{TP}{TP+FN} \in [0,1] \\ Correctness &= \frac{TP}{TP+FP} \in [0,1] \end{aligned} \qquad (1)$$

In Equation 1, *TP*, *FP*, and *FN* are the numbers of true positives, false positives, and false negatives, respectively. They refer to changes in the change map compared to actual changes in the reference. Thus, a *TP* is an entity reported as *changed* (*demolished* or *new*) that is actually changed in the reference. A *FP* is an entity reported as *changed* by an algorithm that has not *changed* in the reference. A *FN* is an entity that was reported as *unchanged* by an algorithm, but is *changed* in the reference. Finally, an entity reported as *unchanged* by an algorithm and also being *unchanged* in the reference is a true negative (*TN*). In this context, the entities to compare can be buildings, which results in per-building quality measures, or pixels in a rasterised version of the change map, which results in per-pixel quality measures. In the cases where it was decided to use a building label image representing the updated map for the evaluation, the rules for classifying an entity as a *TP*, a *FP*, a *FN*, or a *TN* had to be defined in a slightly different way. Any existing (i.e. *unchanged*) building in the reference database is considered a *TN* if a predefined percentage (T_h) of its area is covered with buildings in the new label image. Otherwise, it is considered a *FP*, because it does not have a substantial correspondence in the new label image, which thus indicates a change. A *demolished* building in the reference database is considered a *TP* if the percentage of its area covered by any building in the new label image is smaller than T_h. Otherwise, it is considered to be a *FN*, because the fact that it corresponds to buildings in the new label image indicates that the change detection algorithm has not found this building to have been demolished. A *new* building in the reference database is considered a *TP* if the cover percentage is greater than T_h. Otherwise, it is considered a *FN*. The remaining large areas in the new label image that do not match any of the previous cases correspond to objects wrongly alerted as *new* by the algorithm and thus constitute *FP*s.

3. CHANGE DETECTION APPROACHES

3.1 Method 1 - (Matikainen et al., 2007)

The building detection method of the Finnish Geodetic Institute (FGI) was originally developed to use laser scanning data as primary data. In this study, it is directly applied to input correlation DSM and orthophotos. The method includes the following stages:

1. Pre-classification of DSM height points to separate ground points from above-ground points, using (Terrasolid, 2008).
2. The region-based segmentation of the DSM into homogeneous regions and calculation of various attributes for each segment, using (Definiens, 2008)
3. The classification of the segments into ground and above-ground classes by using the pre-classification.
4. The definition of training segments for buildings and trees on the basis of training data sets.
5. The construction of a classification tree by using the attributes of the training segments (Breiman et al., 1984).
6. The classification of above-ground segments into buildings and trees on the basis of their attributes and the classification tree.
7. A post-processing to correct small, misclassified areas by investigating the size and neighbourhood of the areas.

The output of the method also consists of a building label image representing the new state of the database, which is used for evaluation in this study.

3.2 Method 2 - (Rottensteiner, 2007)

The input data of this method consist of a DSM obtained by LIDAR or stereo-matching techniques. A geocoded NDVI image, the initial building data base, and a Digital Terrain Model (DTM) can optionally be used. If no DTM is available, it is derived from the DSM by hierarchic morphologic filtering. If the initial database is available, it can be used to introduce a bias that favours a classification consistent with the initial data base, because in most scenes only a small percentage of buildings will actually have changed. The workflow of the method consists of three stages. First, a Dempster-Shafer fusion process is carried out on a per-pixel basis and results in a classification of the input data into one of four predefined classes: *buildings*, *trees*, *grass land*, and *bare soil*. Connected components of building pixels are grouped to constitute initial building regions. A second Dempster-Shafer fusion process is then carried out on a per-region basis to eliminate regions corresponding to trees. The third stage of the work flow is the actual change detection process, in which the detected buildings are compared to the existing map. A very detailed change map is generated in this process. The output of the method consists of a building label image representing the new state of the data base and in change maps describing the change status both on a per-pixel and a per-building level (Refer to (Rottensteiner, 2007) for more details). Since the definitions of the classes in the change map do not match those required in Section 2.2, the building label image is used for evaluation in this study.

3.3 Method 3 - (Champion, 2007)

The input data of the method are a DSM, a vegetation mask, computed from CIR orthophotos and NDVI index, and a DTM, automatically derived from the DSM using the algorithm described in (Champion and Boldo, 2006). The workflow consists of 2 stages: in a first step, geometric primitives, extracted from the DSM (2D contours i.e. height discontinuities) or from multiple images (3D segments, computed with (Taillandier and Deriche, 2002)), are collected for each building and matched with primitives derived from the existing vector map. A final decision about acceptance or rejection is then achieved per building. In the second step, the DTM is combined with the DSM to process an above-ground mask. This mask is morphologically compared to the initial building mask (derived from the vector database) and the vegetation mask and new buildings are extracted. The output of the method consists of a change map, in which each building is labelled as *unchanged*, *demolished* or *new*.

4. RESULTS AND DISCUSSION

The evaluation outputs are summarized in Table 1. The completeness and correctness are given for each test area and for each approach, both on a per-building basis and on a per-pixel basis. Values in bold indicates for which methods the best results are achieved. In an optimal system, completeness and correctness are equal to 1: in that case, there is no *FN* (no under-detection) and no *FP* (no overdetection). To be of practical interest, i.e. to consider the system effective and operational, previous works on change detection (Steinnocher and Kressler, 2006), (Mayer et al., 2006) expect a completeness rate close to 1 (typically 0.85) and a high correctness rate (typically 0.7). These recommendations are true for completeness (the new map must be really *new*) but must be modulated for correctness, with respect to the type of change. In our opinion, the effectiveness of a system is mostly related to the amount of work saved for a human operator. That also corresponds to the number of *unchanged* buildings that are correctly detected, because these buildings need no longer be inspected. As a consequence, the correctness rate has to be high during the verification of the database (i.e. for *demolished* buildings). By contrast, a low correctness rate for new buildings is less problematic: without the support of automatic techniques, the entire scene has to be examined by a human operator and therefore, a system that delivers a set of potential new buildings is effective if the true changes are contained in the set and if the number of *FP* s is not overwhelmingly large.

4.1 Marseille Test Area

4.1.1 (Matikainen et al., 2007)

The evaluation of this method is illustrated in Fig. 1-a. The method best operates in term of completeness (0.98). Only two changes are missed by the algorithm and are related to errors in the initial "ground"/"above-ground" classification. Most false alarms are caused by one of two problems, namely the uncertainty of the classification in shadow areas (e.g. Fig. 2-a) and errors in the DSM between buildings that cause a misclassification of street areas as *new* buildings (e.g. Fig. 2-b).

4.1.2 (Rottensteiner, 2007)

The evaluation of this method is illustrated in Fig. 1-b. Overall, the changes are detected correctly. Compared to (Matikainen et al., 2007), there are only three additional *FN*s. Five new buildings are missed by the algorithm (Fig. 2-c), which is caused by errors that occur in the DTM, by complicated topographic features (cliffs). In presence of such features, the DTM, derived from the DSM by hierarchic morphological opening, is less accurate, which predictably limits the extraction of new buildings that is partly based on the difference between the DSM and DTM. Quantisation effects in the DSM (the numerical resolution of height values is restricted to 20 cm) prevented the use of surface roughness as an input parameter for the Dempster-Shafer fusion process, which might have helped to overcome such problems. However, the correctness of the system is acceptable, which implies a limited number of *FP*s. Compared to (Matikainen et al., 2007) and (Champion, 2007), no *FP*s are generated during the detection of new buildings (e.g. Fig. 2-d). *FP*s are mostly caused by small buildings, located in inner yards and in shadow areas: the complexity of the urban scene and the relatively poor quality of the DSM in shadow areas clearly and significantly deteriorate the change detection correctness here.

Method	Completeness		Correctness	
	per building	per pixel	per building	per pixel
Marseille Test Area				
Matikainen	**0.98**	**0.99**	0.54	0.79
Rottensteiner	0.95	0.98	**0.58**	**0.83**
Champion	0.94	0.94	0.45	0.75
Toulouse Test Area				
Rottensteiner	**0.85**	0.90	0.49	0.53
Champion	0.80	**0.95**	**0.55**	**0.85**

Table 1. Completeness and Correctness achieved by the three algorithms for both data sets.

4.1.3 (Champion, 2007)

The evaluation of this method is illustrated in Fig. 1-c. Five *FN*s appear with the method. Two of them (in the north-western corner of the scene - Fig. 2-e) occur during the first stage of the algorithm. They are caused by extracted primitives that are wrongly used to validate demolished buildings. Remaining *FN*s are related to inaccuracies in the processed DTM and occur where topography is particularly difficult. Here again, the overestimation of the terrain height in the DTM prevents the complete extraction of new buildings. Regarding *FP*s, those that occur in the first stage of the algorithm are related to the complexity of the scene: the extraction of pertinent primitives for inner and lower buildings is more difficult and makes the verification more uncertain. Most *FP*s that occur in the second stage are related to building-like structures (walls that are wrongly considered to be new buildings), errors in the vegetation mask (omitted trees) and the same inaccuracies in the correlation DSM (large overestimated areas in narrow streets) that caused errors in the classification of (Matikainen et al., 2007) (Fig. 2-f).

4.1.4 Remark concerning the aerial context

In the context of aerial imagery, there is not a visible predominance of an approach over another one. The three of them perform well in terms of completeness. The differences almost entirely appear in terms of correctness: classification-based methods, i.e. (Matikainen et al., 2007) and (Rottensteiner, 2007), seem to be more efficient than (Champion, 2007) and deliver fewer *FP*s. The per-building correctness rates obtained with the three approaches (0.54, 0.58 and 0.45, respectively) are relatively low, and none of the approaches appears to be a viable basis for a practical solution. However, this consideration must be modulated by the good correctness values (of 0.79, 0.83 and 0.75), that are computed on a per-pixel basis. Since the per-pixel values are directly linked to the area that has been classified correctly or not, these values clearly highlight that the change detection is mostly uncertain for small buildings.

(a) Matikainen

(b) Rottensteiner

(c) Champion

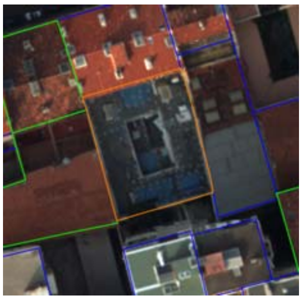
Figure 1. Change Detection Evaluation in Marseille Test Area. In green, *TP* cases; in red, *FN* cases; in orange, *FP* cases; in blue, *TN* cases.

4.2 Toulouse Test Area

4.2.1 (Rottensteiner, 2007)

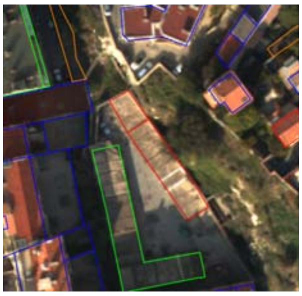
The evaluation of this method is illustrated in Fig. 3-a. Again, major changes are well detected. However, changes affecting small buildings are missed, which results in a high number of *FN*s for small buildings (Fig. 4-a). There are also many *FP*s that are mostly caused by inaccuracies in the DSM. Shadow areas are also systematically overestimated in the DSM, which generates *FP*s during the detection of new buildings. Two very large areas of false alarms appear in the eastern part of the scene (a sports field - Fig. 4-b) and in the north-eastern corner (a parking lot) and are related to classical problems of stereo-matching algorithms, namely repeating patterns (demarcation lines in the sports field, rows of cars on the parking lot) and poor contrast. This entails height variations larger than 4 m in the surface model in areas that are essentially horizontal.

4.2.2 (Champion, 2007): The evaluation of this method is illustrated in Fig. 3-b. Overall, the results are similar to those computed by (Rottensteiner, 2007) and particularly poor for small buildings (Fig. 4-c). The difficulty to extract pertinent primitives for small buildings entails 8 *FN* cases during the detection of new buildings and many *FP*s during the verification of the database (cf. Fig. 4-d for an example).

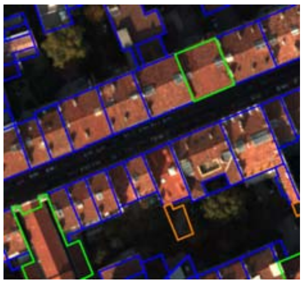


(a) Matikainen: *FP* case

(b) Matikainen: *FP* case

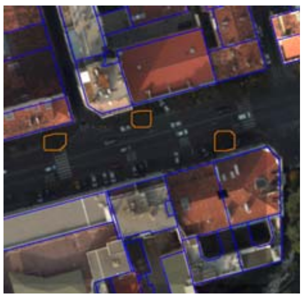


(c) Rottensteiner: *FN* case



(d) Rottensteiner: no *FP* case

(e) Champion: FN cases

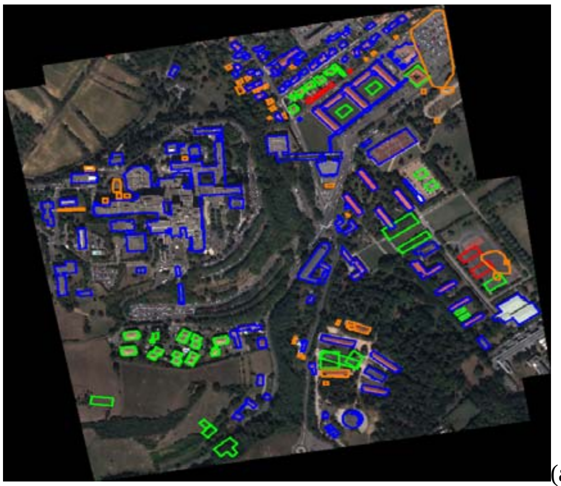


(f) Champion: *FP* cases

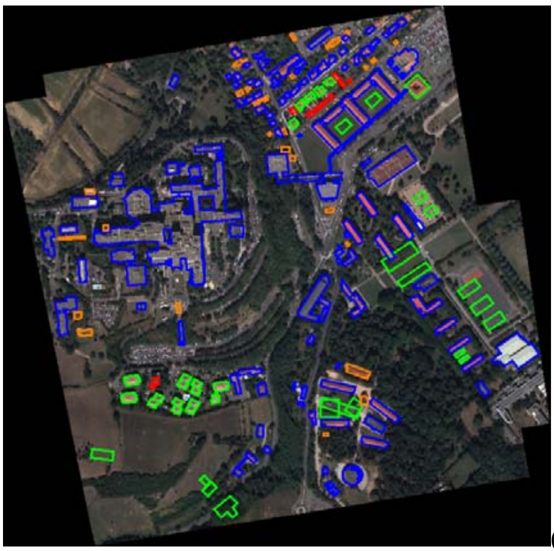
Figure 2. Evaluation details in Marseille. The same colour code as Fig. 1 is used.

4.2.3 Remark concerning the satellite context

The change detection is clearly limited by the resolution of input satellite data in relation to the size of changes to be detected. The completeness and correctness values are (0.85, 0.50) and (0.80, 0.55) for the two methods, respectively. Such values clearly reflect that detecting 2D building changes in a satellite context is too hard a challenge for the current state-of-the-art. This observation is corroborated by the fact that the ground classification performed in (Matikainen et al., 2007) does not give acceptable results for Toulouse when carried out in a fully automatic way. The results may appear to be disappointing. However, the completeness of the two systems both turn out to be very close to the values found in (Mayer et al., 2006) and expected for a system to be operational. Regarding the correctness, the low rate is mostly related to a high number of *FP*s during the detection of new buildings, especially for (Rottensteiner, 2007). As mentioned at the beginning of section 4, it is less problematic as the corresponding *FP* objects would inescapably be checked by a human without the support of automatic techniques. The results that are achieved here clearly demonstrate that it is worth while to carry out research in the satellite context, especially towards the reduction of *FP*s.


(a)

(a) Rottensteiner


(b)

(b) Champion

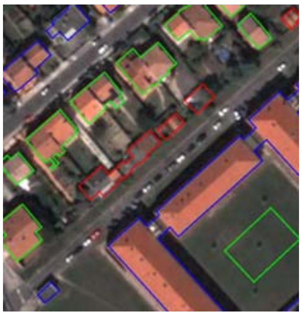
Figure 3. Change Detection Evaluation in Toulouse Test Area (with the same colour code as in Fig. 1).

4.3 Factors affecting the accuracy

In this section, we will try to sum up some preliminary results based on the experiences of this EuroSDR project. We will focus the analysis on the impact of input data on the change detection performance on the one hand, on the features of the method on the other hand, more specifically on the type (geometric/radiometric) of primitives to use.

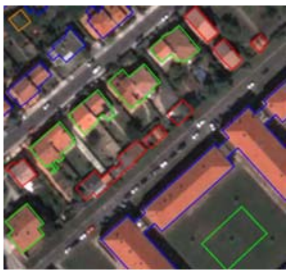
4.3.1 Impact of the DSM

The limiting factor of change detection appears to be the quality of the DSM. The erroneous height values present in the initial DSM between some buildings (i.e. nearby step edges) and the quantization effect observed in both areas and that prevents to exploit surface roughness in the change detection process clearly affect the quality of output change maps. It should be possible to overcome such drawbacks by using LIDAR data, as indicated by the completeness and correctness computed in (Matikainen et al., 2004) and (Rottensteiner, 2007). The results achieved for the EuroSDR test dataset based on LIDAR data have not been evaluated yet but should confirm those results. Improving the performance of an image-based change detection system implies a higher robustness of stereo-matching techniques with respect to shadow areas and a higher preservation of object details, especially step edges.



(a) Rottensteiner: *FN* case

(b) Rottensteiner: *FP* case



(c) Champion: *FN* cases

(d) Champion: *FP* cases

Figure 4. Evaluation details in Toulouse.

4.3.2 Impact of the DTM

As previously highlighted, the extraction of buildings and consequently the performance of the change detection process are the better the more accurate the DTM is. In this study, the morphology-based method used in (Rottensteiner, 2007) and the surface-based method (Champion and Boldo, 2006) both fail in the presence of topographic discontinuities (cliffs). Refined approaches should be considered in the future to better model such terrain features. It must also be noted that in order to keep the process fully automatic, manual corrections were not employed for this study. Manual corrections of difficult points are a normal practice in operational processing and could for instance be used to correct the initial classification into terrain vs. off-terrain points required for DTM generation. This would give a better basis for the later classification stages and improve the final change map.

4.3.3 Impact of the vegetation mask

As previously mentioned, some confusion occurs in this test between trees and buildings. One solution may consist in using more robust criteria to extract vegetation. Classification-based methods, similar to (Trias-Sanz et al., to appear) could also be preferred to NDVI-based methods. Another hint consists in introducing an ontology for objects of interest in the scene (at least for buildings and trees): simple rules such as, "a tree cannot be entirely contained in a building", could easily limit the number of errors, such as the *FN* case produced by (Champion, 2007) in the north-western part of Marseille test (Fig. 2-e).

4.3.4 Impact of the primitives used in the system

In the three approaches, geometric primitives are preferred to radiometric features. In a LIDAR context, such an approach is valid, as the geometry is known to be well described. The image context is more difficult: the 2D contours, extracted in the DSM with (Champion, 2007) are less accurate and the surface roughness computed in (Rottensteiner, 2007) is meaningless. Finding alternative and robust geometric primitives, such as 3D segments (Taillandier and Deriche, 2002), is therefore of high interest. Another solution consists in using radiometric primitives. The use of colour information could also limit the number of *FN*s, such as those produced by (Champion, 2007) in Marseille (Figure 2-f) and could limit the large *FP* areas that occur in Toulouse with (Rottensteiner, 2007). The performance of a change detection system also seems to be closely related to the right combination of radiometric and geometric features.

5. CONCLUSION

Three change detection methods have been tested and compared in this study. The scope has been limited to the imagery context, although two of the three methods, i.e. (Matikainen et al., 2007) and (Rottensteiner, 2007) were not originally designed to deal with imagery. The results presented in this paper demonstrate their high transferability on the one hand and the potential of imagery as an alternative to LIDAR data to detect changes in a 2D building database on the other hand. The results are particularly good, especially in terms of completeness and show the significance to use such interactive techniques in an updating process. The remaining work to be done concerns the reduction of false alarms. The study clearly shows that geometric primitives (height, roughness. . .) that are known to be pertinent in a LIDAR context are less accurate when imagery is used. The improvement of DSM quality is a key point, but other solutions (extraction of new geometric primitives, better modelling of the terrain, and integration of radiometric primitives) are interesting research directions, too. The main focus of the project now is on evaluating more thoroughly the performance of the system with respect to the update status of the building (*unchanged*, *demolished* or *new*) and its size on the one hand, on evaluating the results processed for an area that contains LIDAR data (Lyngby, Denmark) on the other hand. The preliminary visualization of the results on this test area shows a good behaviour of (Matikainen et al., 2007) and (Rottensteiner, 2007). Quantitative evaluation is still necessary to evaluate their performance and the transferability of (Champion, 2007) that is originally built to deal with aerial images. Finally, we plan to expand this test to other methods, namely those described in (Olsen and Knudsen, 2006) and (Katartzis and Sahli, 2008). Beyond scientific results, we hope that this project will be a good opportunity to create a network of interested people both in academia and in the private sector that can speed up the progress in the field of automated building change detection.

ACKNOWLEDGEMENTS

We are grateful to EuroSDR for giving us the opportunity to conduct this test and CNES for providing Pléiades imagery.

REFERENCES

Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J., 1984. Classification and Regression Trees. The Wadsworth Statistics / Probability Series, Wadsworth, Inc., Belmont, California.

Champion, N., 2007. 2D building change detection from high resolution aerial images and correlation Digital Surface Models. In: Int'l Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI-3/W49A (on CD-ROM), pp. 197–202.

Champion, N., Boldo, D., 2006. A robust algorithm for estimating Digital Terrain Models from Digital Surface Models in dense urban areas. In: *Int'l Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI-3, pp. 111–116.

Definiens, 2008. ttp://www.definiens.com/.

Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., 1997. Automated reconstruction of topographic objects from aerial images using vectorized map information. In: *Int'l Archives of Photogrammetry and Remote Sensing*, XXXII, pp. 47–56.

Katartzis, A., Sahli, H., 2008. A stochastic framework for the identification of building rooftops using a single remote sensing image. *IEEE Transactions on Geoscience and Remote Sensing* 46, pp. 259–271.

Matikainen, L., Hyyppä, J., Kaartinen, H., 2004. Automatic detection of changes from laser scanner and aerial image data for updating building maps. In: *Int'l Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXV-B2, pp. 434–439.

Matikainen, L., Kaartinen, K., Hyyppä, J., 2007. Classification tree based building detection from laser scanner and aerial image data. In: *Int'l Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI, pp. 280–287.

Mayer, H., Hinz, S., Bacher, U., Baltsavias, E., 2006. A test of automatic road extraction approaches. In: *Int'l Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI-3, pp. 209–214.

Olsen, B., Knudsen, T., 2006. Building change detection: a case study comparing results from analog and digital imagery. Technical report, KMS (National Survey and Cadastre) and Danish National Space Centre.

Pierrot-Deseilligny, M., Paparoditis, N., 2006. An optimization-based surface reconstruction from Spot5- HRS stereo imagery. In: *Int'l Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI-1/W41, pp. 73–77.

Rottensteiner, F., 2007. Building change detection from Digital Surface Models and multi-spectral images. In: *Int'l Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI-3/W49B (on CD-ROM).

Steinnocher, K., Kressler, F., 2006. Change detection. Technical report, EuroSDR Report.

Taillandier, F., Deriche, R., 2002. A reconstruction of 3D linear primitives from multiple views for urban areas modelisation. In: *Int'l Archives of Photogrammetry and Remote Sensing*, XXXIV-3B, pp. 267–272.

Terrasolid, 2008. http://www.terrasolid.fi/.

Trias-Sanz, R., Stamon, G., Louchet, J., To Appear. Using colour, texture, and hierarchial segmentation for high-resolution remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*.