QUALITY EVALUATION OF 3D CITY BUILDING MODELS WITH AUTOMATIC ERROR DIAGNOSIS

Jean-Christophe Michelin[1,2], Julien Tierny[2], Florence Tupin[2], Clément Mallet[1] and Nicolas Paparoditis[1]

[1] IGN/SR, MATIS, Université Paris Est, 73 avenue de Paris, 94160 Saint-Mande, France,
(firstname.lastname@ign.fr),
[2] Institut Mines-Télécom, Télécom ParisTech, LTCI, 46 rue Barrault, 75634 Paris Cedex 13, France,
(firstname.lastname@telecom-paristech.fr)

Commission III/4

ABSTRACT:

Automatic building modelling allows a cost effective access to 3D semantic information of cities. However, even state-of-the-art algorithms have intrinsic limits and many errors exist in 3D reconstructions, requiring expensive manual corrections. A new approach is proposed in this paper for the automatic diagnosis of 3D building databases in urban areas. A novel error taxonomy which allows a subsequent high-level diagnosis is first proposed. Then, relevant raster and vector features are extracted from very high resolution multi-view images and Digital Surface Models so as that to retrieve such errors. In a supervised way, a set of functions is presented in order to take high-level decisions from these low-level features. Experiments on 355 buildings in an European dense city center with 10 cm airborne images demonstrate the high accuracy on error detection and show promising results.

1 INTRODUCTION

Over the last two decades, three-dimensional city modelling has been one of the main research issues for urban planning and city understanding, with an emphasis on building detection and re-construction. With various levels of details (LoD) (Gröger and Plümer, 2012). State-of-the-art automatic reconstruction algorithms address the LoD2, which proposes a metric representation of 3D facets of the roofs and facades. However their high quality generation is still a challenge in complex urban environments (Rottensteiner et al., 2012). A large number of methodologies was developed with various remote sensing data sources - optical images (Bulatov et al., 2012), lidar (Lafarge and Mallet, 2012), radar (Sportouche et al., 2011) - mainly focused on building extraction from aerial and satellite optical images. The adopted methods heavily depend on the spatial resolution of the input data and no generic method has been developed so far. Regardless of the computed approach, the evaluation of generated 3D building city models has been barely tackled in the literature, apart from measures proposed in the ISPRS benchmark of Rottensteiner et al. (2012), and simple distances between 3D models and ground truth measurements (Macay Moreira et al., 2013). Currently, evaluation can be only detected with (expensive) visual inspection or with (non-generic) intrinsic criteria of reconstruction methods, which are both unsatisfactory. Automatic diagnosis (AD) appears an invaluable input for several purpose:

- **Operational improvement** of 3D building database (DB). One of the main limits of fully automatic reconstruction algorithms is the difficulty to distinguish the correct reconstructed buildings from the others. AD can put aside the correct ones and selects the erroneous reconstructions for more efficient subsequent manual edition.

- **3D change detection**: if a DB is correct at the time t_1, the AD computation with more recent data acquired at time $t_1 + \epsilon$ may provide a knowledge of areas requiring updates (Taneja et al., 2013).

- **Reconstruction algorithm selection**: a large number of reconstruction techniques exists featuring specific advantages and flows. An AD computed on different reconstructions allows to select the best proposal for a given specification and landscape.

Important and very high-level operational errors (e.g., an inner court is missing in a building) cannot be only computed with low-level geometric measures. Therefore the proposed strategy should provide hierarchized indicators for operational issues. Another major challenge of AD is also one of the main operational implementation difficulties: the proposed AD algorithm needs to be fully independent to the (unknown) reconstruction methodology, in order to avoid under/misdetection. Moreover, such a strategy allow a simple upgrade of production lines and opens up perspectives for controlling data for unknown providers.

This paper aims to propose an automatic strategy for the evaluation of any kind of LoD2 3D building city models. Using very high resolution images (10 cm) in a high multi-stereoscopic configuration, we focus on roof modelling and footprint outlining diagnosis, assuming building roof facets are represented as closed polygons. The proposed algorithm is decomposed in two parts. Firstly different set of features are extracted from DSM and multi-view images. Secondly, these features are compared to DB with a one-against-all classification. Our paper exhibits several new contributions:

- A new taxonomy of errors and their most probable sources. Such a taxonomy has been established with operational services of the French National Institute of Geographic and Forest Information, and corresponds to real customer feedbacks;

- A set of low-level geometric features and interpretations of associated measures to propose a high-level building quality diagnosis;

- A novel analysis strategy, based on these measures, proposes a high level identification of problems.
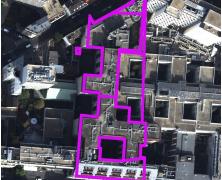
Figure 1: Error taxonomy composed of nine classes, split in three domains For each error, the corresponding roof is superimposed on one of the available aerial images. The proposed color code is specific for each error, and will be employed for the rest of this paper.

2 RELATED WORK

There are two different ways to evaluate 3D models: intrinsic or extrinsic diagnosis. On the one hand, the 3D model generation algorithm is well known (and thereby its drawbacks as well). On the other hand, external data is used to estimate the correctness of the reconstruction. Such data can be very accurate field surveys or very high density terrestrial laser scanning point clouds, similarly to the semi-automatic approach proposed by Helmholz et al. (2013). Oblique aerial images can be a more cost-effective solution. However occlusions management in high density urban cities is one of the major challenges and therefore limit the approach to the 2D building outline diagnosis (Nyaruhuma et al., 2012). Quantitative and qualitative evaluations can involve visual inspection (Durupt and Taillandier, 2006), another reconstructed reference scene (Meidow and Schuster, 2005) or ground truth surveyed measurements (Macay Moreira et al., 2013). These papers highlight the role of the accuracy of the input data (here a DSM) and of the 2D topographic maps used for the reconstruction process. Our paper proposes an extrinsic evaluation to be performed on virtually any LoD2 3D city models.

An evaluation approach based on geometrical errors and segmentation accuracy has to be interpreted with caution (Rottensteiner et al., 2012). A given qualification, without topological clarification, based only on a root mean square (RMS) score is not adequate enough, since a simple 3D mesh (without reconstruction and semantization) would produce the best results. Thus, to diagnose, a building facet criterion is commonly proposed with a topology control strategy (Zhou and Neumann, 2011).

Even if the automatic diagnosis of 3D building remains unsolved in remote sensing, the change detection field has dealt with related problems. Instead, the goal is to detect changes with more recent data than those computed for the DB generation. The automation of a DB diagnosis conventionally based on photo-interpretation has begun more than 15 years ago (Lu et al., 1998). Several methods were proposed with different input data: satellite images (Champion et al., 2010), aerial images (Zhu et al., 2009), aerial lidar (Rutzinger et al., 2010). Other approaches use terrestrial data like panoramical images (Taneja et al., 2013) or terrestrial laser scans (Kang and Lu, 2010). In any case, environmental changes (weather, illumination...) negatively impact the decisions. The proposed solutions rely on the computation of geometrical features: for example, Taneja et al. (2011) test the hypothesis that elements from one image projected on the DB are superimposed with the same elements in other images. If not, a geometry change is assumed to have occurred. Other methods based on image radiometry, e.g., shadows and roof colors (Benedek et al., 2012), exist but they depend on the studied scene types (Olsen, 2004). One can note that, when available, DSM highly improves the classification results (Le Bris and Chehata, 2011). Even if change detection and AD have in common geometrical extraction of features and a classification issue, they differ in several points. Firstly, AD does not assume that the building DB correspond to a ground truth representation. Secondly, data type is different: multi-temporal and more accurate/better resolution data than these used for the DB creation are not accessible. For these reasons, for AD approach, change detection algorithms cannot always be applied straightforwardly.

Recently, to the best of our knowledge, only one paper addressed facet-based self-diagnosis (Boudet et al., 2006). The methodology based on texture and structure analysis allowed a detection of four different roof error classes (false, generalised, acceptable, and correct) but exhibits several limitations:

- **Taxonomy limited** to a particular DB specification: in fact, the error acceptability degree of a DB depends on the user needs. In an operational context, some errors are more important than the others. Thus, an unambiguous estimation in error labelling is expected.

- **Non quantitative assessment**: no confidence score is provided. Therefore, it limits the operational use because no error hierarchy can be proposed. The different kinds of error cannot be established and only one global quality assessment is provided;

- **Limited applications**: this qualification can only be performed as a tool to help a subsequent correction step. The four proposed labels are not high-level enough to propose meta-data.

3 MATERIAL

3.1 Input data

A fully-automatic 3D reconstruction of the city of Paris is available. It was computed with the approach proposed in (Durupt and Taillandier, 2006). In this method, the 2D building footprints are supposed to be known and will be evaluated jointly with the roof facets. To estimate a roof topology, a set of possible simulated planes is computed, then the best model is selected as the one that minimizes a correlation score with a high resolution a 10 cm digital surface model (DSM). The altitude of gutter and roof slopes are estimated with a robust distance minimisation on DSM.

The proposed AD is based on very high resolution aerial images (RGB+IR) which were used to generate the city DB. This data (acquired in 2008) is 10 cm of resolution, with an endlap of 40 % and sidelap of 80 %. This is a high multi-view context: each point of the DB can be projected in images up to 10 times.

3.2 Error taxonomy

In this paper, a study of typical errors of LoD2 city models was performed stemming from requirements of production lines and customers feedbacks of the French National Institute of Geographic and Forest Information, and has led to the creation of a new taxonomy (Figure 1). Nine building errors can be classified in three types:

- **Footprint errors** are caused by an incorrect building outline (errors from input data). Consequently, four sub-errors exist: - erroneous outline: the proposed external footprint does not match with the real footprint;
  - unexisting building: the footprint does not match with any building;
  - missing inner court: the proposed footprint needs an internal polygon but the outline is correct;
  - inaccurate footprint: the footprint matches with the footprint of the ground truth but has a geometrical error, related to a given specification (quantification of accuracy).

- **Reconstruction errors**: inherent to the reconstruction algorithm:
  - under-segmentation: one or more roof ridges are missing;
  - over-segmentation: some parts of roof should not exist;
  - inaccurate roof: a proposed ridge has geometrical errors, related to a given specification;
  - Z translation: the proposed roof is incorrectly positioned in altitude.

- **Vegetation errors** are present when the proposed building is underneath tree canopy. In this case, the AD cannot be reliably computed anymore.

The proposed errors cannot be simply evaluated independently, but this taxonomy definition is easily understandable for an operator. The errors may come from building detection algorithm, wrong eliminations of superstructure in roof calculation, complex buildings out of predefined shapes, vegetation occlusion, etc. Uncertainties in DSM due to occlusion and the regularization step applied during surface reconstruction can imply too soft slopes estimation near altimetric discontinuities and have a significant impact of reconstruction results.

3.3 Error statistics

A ground truth of 355 buildings was created from the DB. For operational issues a maximum of 4 errors were labelled per building. The area was selected due to its large variety of buildings: towers, traditional roofs, complex European buildings, and houses representative of peri-urban areas. One can note that *inaccurate footprint*, *inaccurate roof*, and *Z translation* errors are less important for the DB quality than the others because they do not affect the building representation but only its geometrical accuracy (the proposed topology is correct). One can refer to Figure 2, where an overview with the most prominent error per building is provided: only 21% of labelled buildings do not have any error. In the case of footprint errors, *erroneous outline* (22%) of the total of buildings is the most prevalent error, then *unexisting building* (8%), *inaccurate footprint* (5%), and missing inner court (2%). For reconstruction errors, ratio are also important: *sub-segmentation* (19%), *inaccurate roof* (9%), *over-segmentation* (7%), *Z translation* (4%). Only 2% of buildings have a *vegetation* error: it highly depends on the proportion of vegetation in the study area.

One (or several) sources of error can exist for each building. A study performed in the labelled DB highlights the multiple error sources: erroneous footprint reconstruction (41%), false sub-building segmentation (28%), erroneous estimation superstructure (19%), too complex buildings to be automatically reconstructed (8%), vegetation occlusion -including shrub on balconies- (7%), wrong reconstruction of DSM due to a very high neighbour building (5%).

| Database (top view)	GT error overview
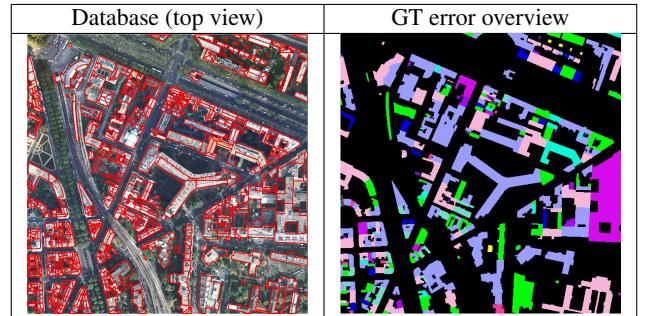
Figure 2: An overview of the Ground Truth (GT): a maximum of 4 labels per element.
Correct ■; Erroneous outline ■; Unexisting building ■;
Missing inner court ■; Inaccurate footprint ■; Under-segmentation ■;
Over-segmentation ■; Roof inaccurate ■; Z translation ■; Vegetation ■.

4 PROPOSED APPROACH

Building qualification depends on an interpretation step and needs to know the expected generalisation degree. To avoid this difficult step, the proposed approach consists in testing for each building the presence of the nine errors contained in the taxonomy (cf Figure 1). Our algorithm can be decomposed into two steps. First, a feature extraction step is performed: for each building a set of vector and raster images are extracted from both aerial images and DSM. Then, a comparison of the features with the DB is proposed to compute a set of descriptors, which finally leads to the classification of each building.

4.1 Feature extraction

Four kinds of features are computed (Figure 3):
- **Image-based 3D segment extraction**: retrieving a very complete set of segments is an important step to verify the building

contours. We use the algorithm proposed by Taillandier and De-riche (2002). For our experience it provides sufficiently accurate results. Using the multi-view context, polygonalized contours are extracted in aerial images and accumulated in the 3D object space with the use of plane sweeping algorithms. Extracted 3D segments are validated with images (quality test: χ^2). These segments correspond to high radiometric discontinuities in images such as building edges, but also road marks, zinc roof battens etc. As illustrated on Figure 3a, the detection is very exhaustive but a filtering step is required to prune the lines which are irrelevant for our purpose (see Section 4.2). The goal of this feature is to detect *over-segmentation*, *inaccurate roof*, *Z translation* and *footprints errors*, and will be noted Σ_{img}.
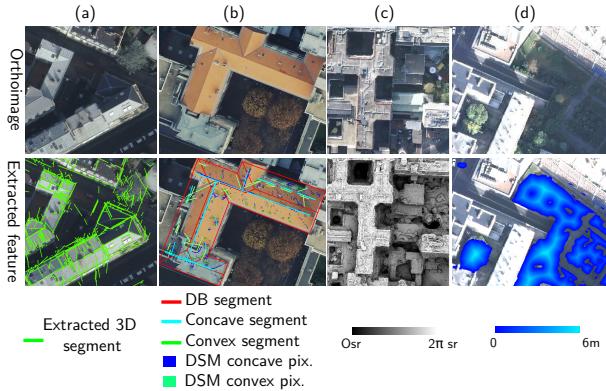

Figure 3: Extraction of feature to detect errors:
a- Image based 3D segment extraction (Σ_{img}), b- Mixed DSM-image based 3D segment extraction (Σ_{mix}), c- Sky Viewshed Angle (SVA), d- Distance based NDVI mask (D_{NDVI}).

• **Mixed DSM-image based 3D segment extraction**. The goal of this feature is to retrieve very robust 3D segments to detect specifically *under-segmentation* errors. Because of the regularisation of the DSM, the resulting surface is too smooth around high discontinuities, what excludes segment extraction near building walls. To locate the DSM discontinuities, a scalar product is computed between robustly extracted normals ($L_{1.2}$ estimator in a spherical neighbourhood of 1 m). One of the keys of the proposed algorithm is the improvement of the robustness by the separation of convex and concave discontinuities. For this purpose, the main direction of each discontinuity pixel is computed using Principal Component Analysis (PCA). The concavity type can be deduced by the analysis of the neighbourhood in the orthogonal direction of the main direction. A Radon transform (Radon, 1986) computed on convex and concave discontinuities allows to obtain 2D line directions. Then, a RANSAC algorithm is performed on the DSM to obtain 3D segments from 2D lines. Because of a high false-detection rate, a filtering step is required. Local gradients in the orthogonal direction of the projected 3D segment in aerial images is computed: segments with non-homogeneous and low gradient are discarded. A non exhaustive but very robust set of segments, which correspond to DSM discontinuities and high gradient values in aerial images, is obtained (Figure 3b). This feature will be noted Σ_{mix}.

• **Sky Viewshed Angle (SVA)** is a DSM-based feature, proposed to detect *missing inner court* errors. This solid angle is computed for each pixel: a 16-direction sample ray tracing allows an admissible computing time with good accuracy. As illustrated on Figure 3c, SVA is a good descriptor for enclosed areas.

• **Distance-based NDVI mask**. The goal of this feature is to detect *vegetation* errors. A true robust NDVI orthoimage is first computed using both DSM and aerial images. In case of ambiguity (i.e. if the average deviation value of the computed point from the median of all images $> 10\%$) no value is computed. A binary threshold of 0.4 is empirically fixed. The obtained mask is propagated using a morphological voting binary hole filter which fills

in holes and cavities by applying a voting operation on each pixel. Finally, an Euclidean distance is computed using the algorithm presented in (Danielsson, 1980). As illustrated on Figure 3d, this feature allows us to assess whether a pixel is near the edge or in the center of a vegetated area. It will be noted D_{NDVI}.

4.2 Error detection

The extracted features and the DB can be compared to detect all errors of the taxonomy. We will note S_{GL}^k the kth segment composing the DB gutter line and respectively S_{R}^k for the kth roof segment.
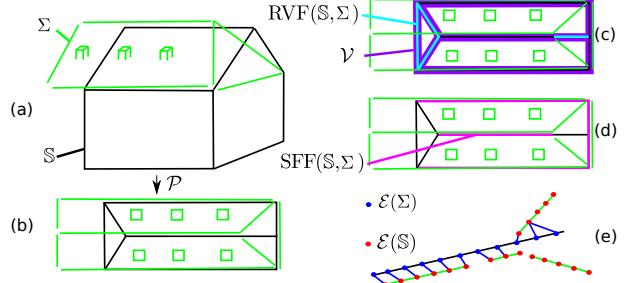

Figure 4: Illustration of the proposed functions for DB error classification.

A template function is defined to compare a set of segments $\mathbb{S} = \{S^1, S^2 ...\}$ from another one $\Sigma = \{s^1, s^2 ...\}$, as illustrated on Figure 4a (l is a segment length). Three points need to be highlighted for the understanding of the proposed computation algorithms:

- A projection $\mathcal{P}$ (on a horizontal or vertical plane) separates planimetic errors from altimetric ones (cf. Figure 4b);

- As explained in Part 4.1, a filtering step based on the segment direction Δ (normalized between 0 and 180°) is computed;

- The neighbourhood $\mathcal{V}$ with a radius of 50 cm centred at of the studied point can be a 2D disk or a 3D sphere.

The sampling step d is fixed to a half pixel resolution (5cm). Four template functions are then defined to characterize the various errors:

• Relative Validation Function (RVF) is defined to compare extracted segments to a subset of DB segments (i.e. gutter lines segment $\mathbb{S}_{\mathrm{GL}}$, roof segments $\mathbb{S}_{\mathrm{R}}$) as:

$$\mathrm{RVF}(\mathbb{S}, \Sigma) = \mathrm{d}\,\frac{\sum\limits_{k=1}^{\mathrm{Card}(\mathbb{S})} \sum\limits_{p=1}^{\lfloor l(\mathbf{S}^k)/\mathrm{d} \rfloor} \delta(\mathbf{S}^k(p), \Sigma)}{\sum\limits_{k=1}^{\mathrm{Card}(\mathbb{S})} l(\mathcal{P}(\mathbf{S}^k))} \quad \text{with:}$$

$$\delta(\mathbf{S}^k(p), \Sigma) = \begin{cases} 0 & \text{if } \mathcal{P}(\mathbf{S}^k(p)) \cap \mathcal{V}(\mathcal{P}(\Sigma)) \neq \varnothing \\ & \text{and } \exists s \in \Sigma \,/\, |\Delta(\mathcal{P}(s)) - \Delta(\mathcal{P}(\mathbf{S}^k(p)))| < \theta \\ 1 & \text{otherwise} \end{cases}$$

(1)

As illustrated on Figure 4c, RVF consists in checking which parts of DB segments are validated by a set of extracted segments projected on a horizontal/vertical plane after a direction thresholding (θ) filtering step. RVF is normalised by the total length of the studying set of segments (i.e. perimeter in case of gutter line). One can note that RVF($\mathbb{S}_{\mathrm{GL}}, \Sigma_{\mathrm{img}}$) checks the *erroneous outline / unexisting building* errors , and RVF($\mathbb{S}_{\mathrm{R}}, \Sigma_{\mathrm{img}}$) the *under-segmentation* error.
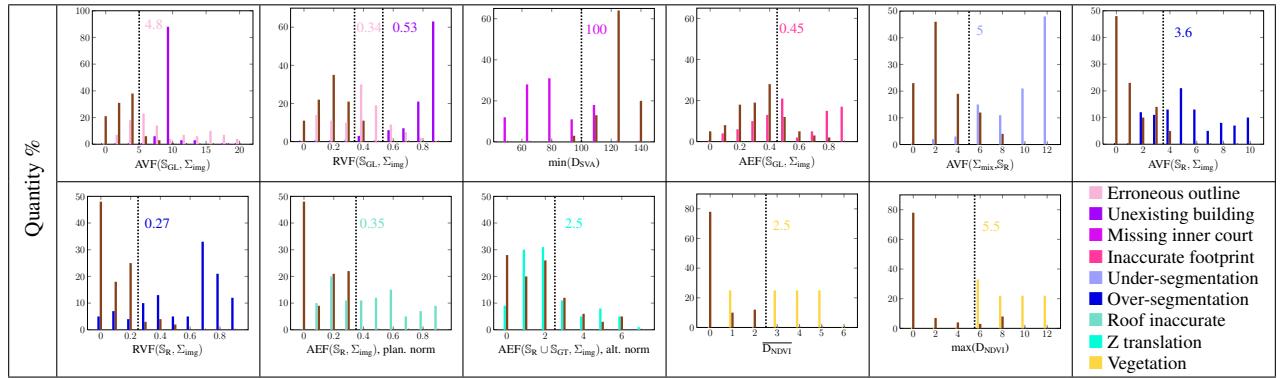
Figure 5: Error detection using one-against-all method (training set). ■: complement of each evaluated error. - -: computed threshold.

In our paper, θ is empirically fixed to 20°, which is a hard threshold to delete non-matching segments.

• **Absolute Validation Function** (AVF) is similar to RVF without normalisation, in order to check large buildings which may be misdescribed with a relative measure. AVF is length of the longest connected set which does not satisfy:

$$\forall k \in [1, \mathrm{Card}(\mathbb{S})] \text{ and } \forall p \in [1; \lfloor l(\mathrm{S}^k)/\mathrm{d}\rfloor] \text{ we have:}$$
$$\mathcal{P}(\mathrm{S}^k(p)) \cap \mathcal{V}(\mathcal{P}(\Sigma)) \neq \varnothing \text{ and } \exists s \in \Sigma \qquad (2)$$
$$\text{such that } |\Delta(\mathcal{P}(s)) - \Delta(\mathcal{P}(\mathrm{S}^k(p)))| < \theta$$

$\mathrm{AVF}(\mathbb{S}_{\mathrm{GL}}, \Sigma_{\mathrm{img}})$ and $\mathrm{AVF}(\mathbb{S}_{\mathrm{R}}, \Sigma_{\mathrm{img}})$ can be applied to complement the number of error detections of the RVF function. Moreover, an over-segmentation detector $\mathrm{AVF}(\Sigma_{\mathrm{mix}}, \mathbb{S}_{\mathrm{R}})$ can be defined with the computation of the mixed DSM-image segments Σ_{mix} which are not validated with DB roof segments $\mathbb{S}_{\mathrm{R}}$.

To detect *inaccurate footprint*, *roof inaccurate*, and *Z translation* we can now define a Spatial Filter Function (SFF) (Figure 4d) which discards non-valid segments, in order to obtain a subset $\mathbb{S}_{\mathrm{SFF}}$ from DB segments:

$$\mathrm{SFF}(\mathbb{S}, \Sigma) = \{\mathbb{S}_{\mathrm{SFF}} \subset \mathbb{S}/\mathrm{AVF}(\mathbb{S}_{\mathrm{SFF}}, \Sigma) = 0\} \qquad (3)$$

• **Spatial Average Error Function** (AEF) is proposed to compute distances between the DB and validated segments:

$$\mathrm{AEF}(\mathbb{S}, T) = \frac{1}{\mathrm{Card}(\mathrm{SFF}(\mathbb{S}, T))} \sum_{S_{\mathrm{SFF}} \in \mathrm{SFF}(\mathbb{S}, T)} \mathcal{Q}(S_{\mathrm{SFF}}, T) \qquad (4)$$

$\mathcal{Q}$ is a pseudo-distance, as illustrated in Figure 4e, between one segment T and a set of segments $\mathbb{S}$. It is defined with a segment sampling $\mathcal{E}$, as follows:

$$\mathcal{Q}(\Sigma, \mathbb{S}) = \frac{\sum_{k=1}^{\mathrm{Card}(\mathcal{E}(\Sigma))} \min(\gamma(\mathcal{E}(\Sigma)_k, \mathcal{E}(\mathbb{S})))}{\mathrm{Card}(\mathcal{E}(\Sigma))} \qquad (5)$$

$\gamma(p, \mathrm{C})$ a set of Euclidean distances between a point p and a point cloud C. This pseudo-distance is based on the minimum closest point. If γ is planimetric, $\mathrm{AEF}(\mathbb{S}_{\mathrm{GL}}, \Sigma_{\mathrm{img}})$ can provide a quality score for *inaccurate building* errors, and respectively $\mathrm{AEF}(\mathbb{S}_{\mathrm{R}}, \Sigma_{\mathrm{img}})$ for *inaccurate footprint*. If γ is altimetric, *Z translation* can be quantified with $\mathrm{AEF}(\mathbb{S}_{\mathrm{R}} \cup \mathbb{S}_{\mathrm{GT}}, \Sigma_{\mathrm{img}})$.

• *Vegetation* and *missing inner court* errors cannot be diagnosis with DB segments. Nevertheless, functions on raster features like the average $\overline{\mathrm{D}_{\mathrm{NDVI}}}$ and $\max(\mathrm{D}_{\mathrm{NDVI}})$ (for large buildings) detect *vegetation* errors. In addition, $\min(\mathrm{D}_{\mathrm{SVA}})$ is a *missing inner court* indicator.

5 RESULTS

Half of the labelled buildings is chosen as a training set to study the discrimination potential of the proposed functions. The distribution of the errors in the DB was carefully preserved in the training set. Figure 5 illustrates the error distribution for each function in a one-against-all approach. For each error, the proposed function(s) is(are) computed on the training set: a one-against-all histogram is represented with computed score. Therefore, it demonstrates how the proposed functions are relevant. The less the histograms overlap, the more discriminated error measure will be (error and complement error). The threshold is fixed to the best value which separates the two distributions according to inter/intra classes separation criterion. One can note that the proposed functions are satisfactory to discriminate the coarsest errors of the training data. These errors are the most important because they drastically affect the building representation. One can note that of the most frequent errors (cf. Section 3.3) are satisfyingly discriminated.

A simple binary one-against-all classification is then performed on the second part of labelled DB (test set). Thresholds (i.e. the point which maximizes separations) are selected thanks to the study of histograms with inter/intra classes variances criterion (Figure 5).

Error type	%			
	TP	**FP**		
	FN	**TN**		
Erroneous outline	73	27		
	13	87		
Unexisting building	82	18		
	9	91		
Missing inner court	79	21		
	16	84		
Inaccurate footprint	61	39		
	43	57		
Under-segmentation	78	22		
	24	76		
Over-segmentation	68	32		
	16	74		
Inaccurate Roof	64	36		
	34	66		
Z translation	71	29		
	63	37		
Vegetation	100	0		
	22	78		

(a)
(b)
(c)

■: training set ■: detected ■: not detected
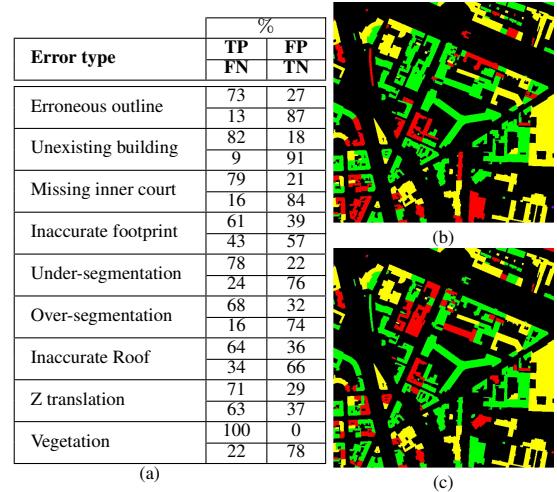
Figure 6: (a) Confusion matrix; (b) Results for the proposed GT overview; (c) Results for the *erroneous outline* error.

As illustrated on Figure 6a and b, results are satisfactory with an average TP rate of 75%. 100% of buildings in vegetation are correctly classified. *Erroneous outline* errors are detected with a rate of 82% and the confusion matrix of *unexisting building* is almost an identity matrix, which implies a very confident detection. In Figure 6c, the proposed descriptor for *erroneous outline* error is very discriminative. One can note that this error detection is a typical operational issue: this error needs absolutely to be detected and corrected (source of others errors).

Missing court errors are correctly retrieved. However, this performance could be improved without the errors due to DSM inaccuracies. *Over* and *under-segmentation* errors are also well determined. Classification of *inaccurate footprint*, *inaccurate roof* and *Z translation* is less efficient. This problem can be explained by several reasons. First, these errors are the finest and consequently the most difficult to detect. Secondly, these error detections are linked to the other ones: *Z translation* is often correlated with *over-segmentation* and the proposed function fails by attempting the separation.

One can note that TP, FP, FN, TN rates of training and test sets are almost similar ($\pm$ 5%) for coarser errors, which implies that the proposed approach exhibits good generalization properties.

6   CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a novel solution for 3D building city models automatic quality diagnosis. It is based on a new error taxonomy and features extracted from very high resolution multiview images. To solve this complex problem we propose to detect different high-level errors. Therefore, a study was performed to know the different type of errors in building DB. The main concern of this paper was to detect several errors for each building, that allows a larger number of applications, instead of a pure binary result. Quantitative results are acceptable, with satisfactory true positive rates (>70%) for the most important errors. Furthermore results can be improved with some recent approaches providing additional and helpful hints such as superstructures (Brédif et al., 2007). The major limitation of this approach is the difficulty to detect finest errors as *inaccurate roof* and *inaccurate footprint* because they are mixed with coarser errors. Another limitation is that only linear low-level primitives are extracted. However, more advanced primitives (curves or key-points) may be inserted in the feature extraction computation to apply the proposed approach to other types of landscapes. Future works will focus on the increase of the size of the testing database to test the generalisation of the proposed approach to a larger area. Finally we would like to propose a more sophisticated classification to significantly improve the results, and to handle uncertainties in the labelling process.

References

Benedek, C., Descombes, X. and Zerubia., J., 2012. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. IEEE TPAMI 34(1), pp. 33–50.

Boudet, L., Paparoditis, N., Jung, F., Martinoty, G. and Pierrot-Deseilligny, M., 2006. A supervised classification approach towards quality self-diagnosis of 3D building models using digital aerial imagery. IAPRS vol. 36(3), pp. 136-141.

Brédif, M., Boldo, D., Pierrot-Deseilligny, M. and Maitre, H., 2007. 3D Building Reconstruction with Parametric Roof Superstructures. ICIP, vol. 2 pp. 537–540.

Bulatov, D., Rottensteiner, F. and Karsten, S., 2012. Context-based urban terrain reconstruction from images and videos. IAPRS vol.1(3) 1(2), pp. 185–190.

Champion, N., Boldo, D., Pierrot-Deseilligny, M. and Stamon, G., 2010. 2D building change detection from high resolution satellite imagery: A two-step hierarchical method based on 3D invariant primitives. Pattern Recognition Letters vol. 31(10), pp. 1138-1147.

Danielsson, P., 1980. Euclidian distance mapping. Computer graphics and image processing vol. 14.

Durupt, M. and Taillandier, F., 2006. Automatic building reconstruction from a digital elevation model and cadastral data: An operational approach. IAPRS vol. 36(3).

Gröger, G. and Plümer, L., 2012. CityGML Interoperable semantic 3D city models. ISPRS Journal 71, pp. 12 – 33.

Helmholz, P., Beldon, D. and Moncrieff, S., 2013. Approach for the semi-automatic verification of 3D buildings models. IAPRS vol. 40(1/W1), pp. 121-126.

Kang, Z. and Lu, Z., 2010. The change detection of building models using epochs of terrestrial point clouds. IAPRS vol. 38(8), pp. 231-236.

Lafarge, F. and Mallet, C., 2012. Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation. IJCV vol. 99(1), pp. 69-85.

Le Bris, A. and Chehata, N., 2011. Change detection in a topographic building database using submetric satellite images. IAPRS vol. 38(3/W22), pp. 25-30.

Lu, W., Doihara, T. and Matsumoto, Y., 1998. Detection of building changes by integration of aerial imageries and digital maps. IAPRS vol. 32, pp. 314-317.

Macay Moreira, J., Nex, F., Remondino, F. and Jim, N., 2013. From DSM to 3D building models : quantitative evaluation. IAPRS vol. 40(1), pp. 213-219.

Meidow, J. and Schuster, H.-F., 2005. Voxel-based quality evaluation of photogrammetic building acquisitions. IAPRS vol.36(3/W24), pp. 117-122.

Nyaruhuma, A. P., Gerke, M., Vosselman, G. and Mtalo, E. G., 2012. Verification of 2D building outlines using oblique airborne images. ISPRS Journal 71, pp. 62 – 75.

Olsen, B., 2004. Automatic change detection for validation of digital map databases. IAPRS vol. 34(B2), pp. 564-574.

Radon, J., 1986. On the Determination of Functions from Their Integral Values along Certain Manifolds. IEEE TIM vol. 5(4), pp. 170-176.

Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S. and Breitkopf, U., 2012. The ISPRS benchmark on urban object classification and 3D building reconstruction. IAPRS vol. 1(3), pp. 293-298.

Rutzinger, M., Rüf, B., Vetter, M. and Höfle, B., 2010. Change detection of building footprints from airborne laser scanning acquired in short time intervals. IAPRS vol. 38(7b), pp. 475-480.

Sportouche, H., Tupin, F. and Denise, L., 2011. Extraction and three-dimensional reconstruction of isolated buildings in urban scenes from high-resolution optical and SAR spaceborne images. IEEE TGRS vol.49 (10), pp. 3932-3946.

Taillandier, F. and Deriche, R., 2002. Reconstruction of 3D linear primitives from multiple views for urban areas modelisation. IAPRS 34(3A), pp. 267–272.

Taneja, A., Ballan, L. and Pollefeys, M., 2011. Image based detection of geometric changes in urban environments. ICCV'11.

Taneja, A., Ballan, L. and Pollefeys, M., 2013. City-scale change detection in cadastral 3D models using images. CVPR'13.

Zhou, Q.-Y. and Neumann, U., 2011. 2.5D building modeling with topology control. CVPR'11 pp. 2489–2496.

Zhu, L., Shimamura, H., Tachibana, K., Liu, Z. and Gong, P., 2009. Building change detection using aerial images and existing 3D data. In: Joint Urban Remote Sensing Event.