Extraction and Three-Dimensional Reconstruction of Isolated Buildings in Urban Scenes From High-Resolution Optical and SAR Spaceborne Images

Hélène Sportouche, *Member, IEEE*, Florence Tupin, *Senior Member, IEEE*, and Léonard Denise, *Member, IEEE*

***Abstract*—In this paper, we propose a new complete semi-automatic processing chain able to provide, from a couple of high-resolution optical and synthetic aperture radar (SAR) images, a simple 3-D reconstruction of buildings in urban scenes. A sequence of processing, exploring the complementarities of both optical and SAR data, is developed for building reconstruction. The chain is decomposed into the main following steps: First, potential building footprints are extracted from the monoscopic optical image through global detection followed by a boundary refinement. Then, the optical footprints are projected and registered into SAR data to get a fine superposition between optical and SAR homologous ground features. Finally, the last step, based on the optimization of two SAR criteria, is performed to deal with building validation and height retrieval. Each of these steps is methodologically described and applied on scenes of interest on Quickbird and TerraSAR-X images. A qualification of each reconstructed building by a score of confidence is then proposed. Good results of building detection are obtained, and relevant height estimations are retrieved.**

***Index Terms*—Building detection, data fusion, footprint extraction, height estimation, high-resolution (HR) images, Quickbird, synthetic aperture radar (SAR), TerraSAR-X, urban scenes, 3-D reconstruction.**

I. Introduction

DURING THE last years, a new generation of satellites (TerraSAR-X, CosmoSkyMed, Quickbird, Ikonos, etc.) providing a large amount of optical and synthetic aperture radar (SAR) data, with metric or submetric resolution, has been born. New approaches, exploring the high detail level characterizing high-resolution (HR) images, have been developed in order to achieve a better scene analysis and understanding in urban or semi-urban scenes. For instance, new methods have been proposed for the detection of various structures and objects (such as buildings or houses [1], [2], roads [3], and bridges [4]) and for the identification of typical regions (such as vegetation, water areas, and shadows [5], [6]).

In particular, challenges are born in the domain of building boundary extraction and building height retrieval for the 3-D reconstruction of buildings in urban areas [7]–[9]. Several techniques, based on optical and SAR imagery, have been developed to deal with this problem and have proved to provide efficient results.

In the optical field, methods for accurate building detection and reconstruction have been presented in a monoscopic context [10]–[12] or in a stereoscopic context (with two or more input images) [13]. For building detection, these methods mostly use different criteria, giving clues about building presence (such as shadow adjacency [1], [14]) and making easier the extraction of building boundaries (such as geometric constraints [1], [15] or morphological and spectral properties [14], [16], [17]). For building height estimation, different kinds of measures have been proposed depending on the input data (measure of the shadow length [18] or parallax in monoscopy and measure of the disparity between homologous pixels or primitives in stereoscopy [19]). Satisfying and relevant results have been generally obtained from stereoscopy. Other methods based on the joint use of optical images and auxiliary input data, such as approximate digital elevation model [20] or cadastral maps [21], [22], have also been performed. Nevertheless, some limitations can be pointed out. First, the usefulness of optical images highly depends on weather conditions. Second, the auxiliary data or the specific configurations (need of multiple images), required as input, are not always available in operational conditions.

In the SAR field, some well-known techniques for building recognition and height estimation have been proposed in a monoscopic [23]–[25], radargrammetric [26], or interferometric [8], [27], [28] framework and also in the context of multi-aspect SAR data [29], [30]. When only a single SAR image is available, classical measure-based methods, using layover [31] or shadow length [23], have been applied for height retrieval. More recently, new statistical [32], electromagnetic, [33]–[35] or simulation-based [36], [37] approaches, founded

Manuscript received July 16, 2010; revised January 14, 2011; accepted February 6, 2011. Date of publication May 11, 2011; date of current version September 28, 2011.

H. Sportouche and F. Tupin are with the Département Traitement du Signal et des Images, Institut Telecom, Telecom ParisTech, Laboratoire Traitement et Communication de l'Information UMR 5141-Centre National de la Recherche Scientifique, 75013 Paris, France (e-mail: helene.sportouche@telecom-paristech.fr; florence.tupin@telecom-paristech.fr).

L. Denise is with the Département IMINT, Thales Communications, 91883 Massy, France (e-mail: leonard.denise@fr.thalesgroup.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2011.2132727

on the presence of building characteristic areas, have proved to provide satisfying results for building reconstruction. In SAR interferometry (InSAR) or radargrammetry, new Markovian [8] or hypothesis-based methods [38] have been proposed and have given good results. Nevertheless, because of the scene complexity in dense urban areas and because of the difficulties inherent to SAR images (speckle noise, geometric deformations, etc.), these results still remain limited (noisy or incomplete) [39].

In a data fusion context, some techniques for urban building detection and height retrieval, combining optical and SAR information, have been recently explored [40]–[43]. Different configurations of input data have been studied and have generated several appropriate methodologies.

When a couple of radargrammetric images [44], [45] or InSAR images [46] and a single optical image are available, the configuration is particularly well adapted to the extraction of 3-D height information from SAR data and to the deduction of building footprint shape from optical data.

When only a couple of a single SAR image and a single optical image is available, the task remains difficult and only few works have been published on this subject. In such a case, intervention of an operator or addition of some auxiliary input data is often required. Moreover, the different steps that collaborate to the 3-D reconstruction of buildings (identification, contour extraction, and height retrieval) are not always all handled. In [47], a new process for the detection of building outlines, based on the fusion of optical and SAR features, is presented. In [48], authors propose to detect new buildings for vector database updating, by performing building identification through the accumulation of evidences issued from both optical and SAR data (optical shadows, optical lines and right angles for building edges, vegetation index, SAR lines, etc.). In [36] and [49], the developed approach focuses on the step of height estimation according to a "*hypothesis generation–rendering–matching*" procedure, while the parameters defining the rectangular building footprints are issued from a geographic information system database [36] or from a manual extraction on the optical image [49].

In such a frame, new approaches have to be developed to solve the problem of matching and registration between optical and SAR homologous features. These approaches have to be adapted to the kind of features characterizing this spatial HR [50], [51].

Recently, authors have been interested in the problem of building change detection in this optical–SAR data fusion framework [48], [49], [52].

In this paper, we propose a new complete semi-automatic processing chain, providing a simple 3-D reconstruction of buildings in urban or semi-urban scenes from HR optical and SAR imagery. The process takes only as input one HR optical image and one HR SAR image of the same area, with a digital terrain model (DTM) on this area. The following constraints have to be satisfied: 1) The chain has to be able to work in operational conditions, which implies the capacity to deal with large areas (in a relatively short time) and with limited intervention of an operator (semi-automatic process); 2) the different steps composing the chain have to provide the planimetric and altimetric information, required for a simple and relevant 3-D reconstruction of buildings; and 3) confidence about the quality of the resulting building reconstruction has to be proposed.

The proposed chain can be decomposed into three main steps:

1) the detection of potential buildings and the extraction of building boundaries from the monoscopic optical image;
2) the projection and the registration of potential building footprints from optical data into SAR data;
3) the height estimation of potential buildings and the validation of building presence from the SAR image.

In the first time, we show that some planimetric information about potential footprint location and dimensions can be provided from the optical data, by performing a process based on geometric, radiometric, and morphological considerations. In the second time, we demonstrate that the introduction of the SAR data allows one to get an altimetric information, by combining a statistical criterion and a radiometric criterion.

The whole sequence of proceedings is applied to a couple of Quickbird and TerraSAR-X images on an urban area (Marseille, France), and the different intermediate results are illustrated and analyzed on two scenes of interest throughout this paper.

This paper is structured as follows: In Section II, the complete process flow is presented, and the different steps composing the chain are briefly exposed. In Sections III, IV, and V, the three main steps of processing are consecutively developed. In Section VI, a discussion about the operational aspects of the chain (parameter setting and computing time) is proposed. In Section VII, conclusions about the proposed approach are drawn, and some perspectives are envisaged.

II. Presentation of the Processing Chain

A. Description of the Complete Process Flow

The proposed sequence of methods aims to explore the complementarities provided by optical and SAR data. The following two strategic choices are made.

1) First, given the complexity characterizing SAR urban scenes at a metric resolution, it is difficult to detect building presence directly from the single SAR image. We propose thus to extract information about potential building location from the optical image.
2) Second, the lateral viewing of SAR sensors implies geometric deformations (like layover and shadow) that can be used to retrieve 3-D information. We propose thus to estimate the building height from the SAR image.

The complete chart flow of the process is shown in Fig. 1. Each of the main steps composing the chain is described in detail in a dedicated section of this paper. In Section III, the step of building boundary extraction from the optical image is presented; its output is a map giving the footprint boundary location of potential buildings in the optical image referential. In Section IV, the step of footprint projection and registration from the optical image into the SAR one is presented; its output is a map relative to the footprint boundary location of potential optical buildings, projected and registered in the SAR image referential. In Section V, the step of height retrieval

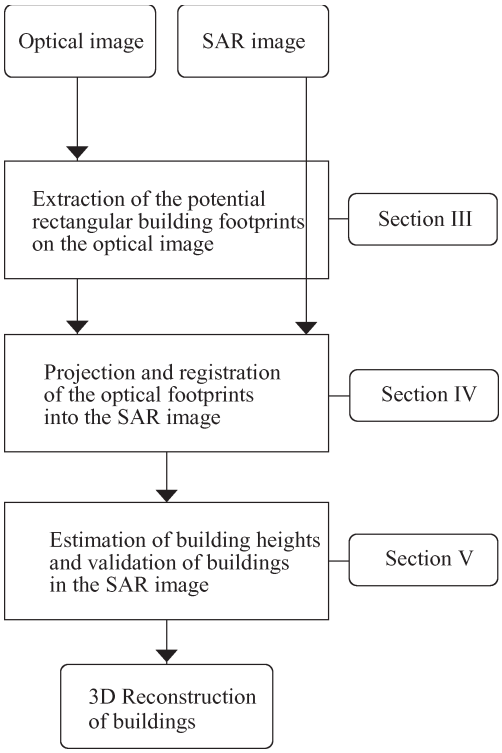


Fig. 1. Chart flow of the complete chain.

and building validation is presented; its output is a map giving the retrieved heights for validated buildings in the SAR image referential.

The set of provided maps is finally used to realize the 3-D reconstruction of buildings in the scene.

B. Hypothesis of Work

We assume that our 3-D building model is a simple parallelepiped (vertical frontages and horizontal flat roof), described by two kinds of parameters: its planimetric rectangular boundary, defined by five parameters (the width, the length, the orientation, and the position of the rectangle center in abscissa and in ordinate), and its height.

We also assume that our optical image has been acquired with a look angle not too far from the nadir. This makes the step of building detection easier, as we do not have to face the difficulty due to frontages. In consequence, the location of the building footprints and the location of the extracted building roof boundaries are the same in the optical image referential.

C. Description of the Data Set

A data set consisting of a Quickbird image (panchromatic mode; resolution of 0.68 m; quasi-nadir look angle) and a TerraSAR-X image (high spotlight mode, resolution of 1.1 m in ground range and 1.1 m in azimuth, incidence of about 32°) on the same area in Marseille (France) is available.

We propose to illustrate the different results, obtained at each step of the chain, on two scenes selected for their variety.

Figs. 2 and 3 show the two scenes of interest studied throughout this paper. The first scene, located in a semi-urban area, is composed of rectangular industrial buildings, comparable with low sheds and relatively far from each other. The second scene, located in an urban area, is composed of rectangular residential buildings, relatively high and very close to each other.

III. Building Footprint Extraction From Monoscopic Optical Image

For the building boundary extraction on the optical image, a two-phase process, providing first a global coarse map and then a refined boundary map, is exposed.

A. Phase 1: Generation of Individual Windows of Interest

In the first phase, a region-based approach is proposed in order to detect individual rectangular areas, likely to be individual buildings, allowing to coarsely generate rectangular boxes called windows of interest.

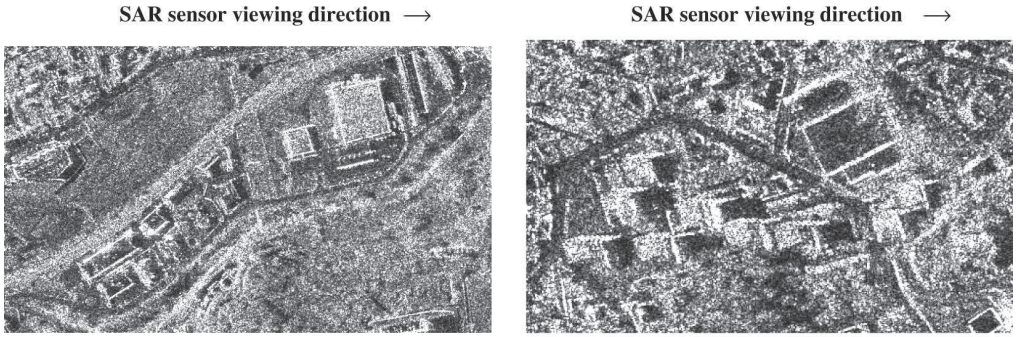
During this phase, a morphological tool, the differential morphological profile (DMP), is used. The concept of DMP, introduced in [53], has already been used for some applications such as classification [5], [6] and change detection [54] in urban areas. The DMP gives information about the size and the contrast of an object. It is built by using opening and closing operators by reconstruction, as these operators permit to preserve structure shapes in the DMP. A multi-scale approach with different structuring elements is adopted to explore a large range of potential object sizes. When a structuring element reaches the characteristic size of an object in the DMP, all its belonging pixels receive the same gray level value (the one from the brighter or darker surrounding region), which induces a peak in the DMP, corresponding to the local contrast. Openings and closings affect, respectively, the structures that are brighter or darker than their surrounding. Objects are thus progressively removed at different levels of the DMP, and a stack of simplified images at different scales is created. In this paper, we take advantage of the properties of image simplification and shape preservation to introduce a geometric criterion based on a prior rectangular building shape. Indeed, as each building appears on the optical image as a rectangular footprint area usually including details (corresponding to superstructures like chimneys or parapets), there is a characteristic level in the DMP, for which all these detail subareas are removed. At this specific level, the building is reduced to a simple rectangular homogeneous area, which can be submitted to a geometrical test.

The proposed method can be decomposed as follows.

1) Preprocessing: First, the shadows, which are clearly visible on the optical data, are extracted by thresholding on the gray levels. Second, the DMP of the image is built and provides a stack of images. Third, the extracted shadow areas are "masked" on the DMP, which means that all pixels included in these shadow areas are replaced by zero values in the different images of the stack, in order to make the building identification easier.

(a) (b)

Fig. 2. Scenes of interest on the optical image (Quickbird sensor; DigitalGlobe; mode: panchromatic; resolution: 0.68 m; area: Marseille, France). (a) Scene 1. (b) Scene 2.



(a) (b)

Fig. 3. Scenes of interest on the SAR image (TerraSAR-X sensor; Infoterra; mode: high spotlight; resolution: 1.1 m; area: Marseille, France). (a) Scene 1. (b) Scene 2.

2) Hierarchical analysis: A test area is defined as a radiometrically homogeneous area, composed of a set of pixels connected with each other and with exactly the same gray level value (connected component).

   The following processing is iteratively applied on the different images in the stack, from the ones corresponding to the larger structuring element to the ones corresponding to the smaller one. For each test area present in the considered image and corresponding to a non-zero gray level local value, a rectangular-oriented bounding box is built using axis of inertia of the area. An overlapping coefficient, characterizing the geometrical adequation between a rectangle of reference and the test area, is computed. This coefficient is defined as the percent of common surface between both surfaces. If it is higher than a threshold, the generated rectangular box is validated as a potential window of interest and the corresponding test area is "masked" for the next iterations, which means that all its pixels are put to zero in the images present in the inferior levels of the DMP. This prevents us from testing the different detail subareas inside the building. As a consequence, for the same real footprint, only one window of interest can be generated.
3) Post-processing: A thresholding on the minimal surface size of the validated windows is realized to delete too small windows.

The obtained coarse map, delineating rectangular windows of interest, is used as input for the next phase.

The different parameters used in this phase of potential window detection are the following:

1) the threshold T_{shadows} that can be locally learnt on shadow areas for a given optical sensor and for given conditions of luminance;
2) the parameters used for DMP construction: the number N_{DMP} of images in the stack and the minimal and maximal sizes $(\epsilon_{\min}, \epsilon_{\max})$ of the structuring element; they depend on the adopted compromise between simplification level in the DMP and computing time;
3) the threshold $T_{\mathrm{overlapping}}$ on the minimal overlapping coefficient that is set to 0.75 by the operator;
4) the minimal surface $S_{\min}$, length $L_{\min}$, and width $W_{\min}$ used in post-processing and deduced at a given resolution from the kind of building size we are looking for.

B. Phase 2: Extraction of Refined Potential Building Footprints

Given that the previous windows of interest have been generated from axis of inertia of rectangular areas, the geometric parameters defining these windows are not always accurate enough. A second phase is thus needed for the refinement of footprint location.

In this phase, a complementary contour-based approach is adopted to ensure a fine superposition between building boundaries and radiometric discontinuities.

We are here locally working around each rectangular window of interest. We try to refine the orientation, dimension, and

position of the rectangular footprint, initialized by the window outline.

First, for each window, a bounding box is defined, and a local Hough accumulator is computed from the thresholded image of Sobel gradient. The principal direction of the building is obtained by selecting in the accumulator the couple (ρ, θ) that induces the larger accumulation value (where ρ and θ correspond, respectively, to the distance coordinate and to the angle coordinate in a polar image referential). The initial rectangle R^0 is then rotated in order to have its main orientation fits with the found direction.

Second, small variations in dimensions and position are envisaged on this new rectangle R^1. A double loop, testing simultaneously a set of possible edge elongations and center translations, is performed on the rectangle R^1. It permits to generate an exhaustive set of test rectangles R^k, defined by

$$R^k = \left(T_{(a,b)} \circ E_c^{\text{length}} \circ E_d^{\text{width}}\right)(R^1) \tag{1}$$

where $T_{(a,b)}$ corresponds to the translation operator according to the vector (a, b), E_c^{length} corresponds to the elongation operator of ratio c applied to both largest edges, and E_d^{width} corresponds to the elongation operator of ratio d applied to both shortest edges.

A stage of score maximization allows to find the optimal rectangle R^*

$$R^* = \arg\max_{R^k} \left(\sum_{(i,j) \in C_{R^k}} \frac{\delta_{(i,j)}}{P_{R^k}} \right) \tag{2}$$

where $\delta_{(i,j)}$ corresponds to the radiometry (0 or 1) of the pixel (i, j) on the thresholded image of Sobel gradient, C_{R^k} designates the contour of R^k, and P_{R^k} corresponds to the perimeter of R^k.

The induced rectangular footprint is finally conserved if the associated maximal score is higher than a threshold T_{score}, assuring a certain level of reliability.

The different parameters used in this phase of footprint refinement are the following:

1) the threshold T_{Sobel} that can be locally learnt on typical building discontinuity areas for a given optical sensor;
2) the maximal norms $a_{\max}$, $b_{\max}$, $c_{\max}$, and $d_{\max}$ of the vectors of translations and elongations that are empirically fixed to a few pixels;
3) the threshold T_{score} set to 0.5 by the operator.

C. Results on the Studied Scenes

The step of building footprint extraction is applied on real data for both studied scenes.

Table I indicates the parameter values used during this step.

Fig. 4 presents the two boundary maps indicating the precise location of the potential footprints in the optical image referential.

After the complete step, nine potential footprints are proposed on scene 1 and eight on scene 2. Among these ones, seven are good detections and two are false alarms on scene 1, and seven are good detections and one is false alarm on scene 2.

TABLE I
PARAMETERS USED FOR BUILDING FOOTPRINT EXTRACTION

	Scene 1	Scene 2
$T_{shadows}$	210	150
N_{DMP}	24	24
$\epsilon_{min}(pixels)$	5	5
$\epsilon_{max}(pixels)$	120	120
$T_{overlapping}$	0.75	0.75
$S_{min}(pixels^2)$	200	500
$L_{min}(pixels)$	20	-
$W_{min}(pixels)$	20	-
T_{Sobel}	10	15
$a_{max}(pixels)$	5	5
$b_{max}(pixels)$	5	5
$c_{max}(pixels)$	7	7
$d_{max}(pixels)$	7	7
T_{score}	0.5	0.5

The radiometric thresholds ($T_{shadows}$ and T_{Sobel}) are given for optical data coded on 16 bits (65536 grey levels).
These data are not calibrated and are directly computed from the radiometrically corrected image pixels (counts) given in standard Quickbird data.
To obtain the associated calibrated values in top-of-atmosphere band-integrated radiance, these corrected values have to be multiplied by the calibration factor K_{band} (here, in P band: $K_{band} = 6.4476 \times 10^{-2}\ W \cdot m^{-2} \cdot sr^{-1} \cdot count^{-1}$).
In this way, the converted values may be used for data from other sensors, which improves the portability of the algorithm.

Concerning good detections, accurate results of location are provided: the boundary parameters, obtained for each building, seem quite satisfying; the deduced coordinates of building corners will be used as a good starting point for the fusion with SAR data.

Concerning false detections, a false alarm rate of 18% is globally obtained on both scenes. These false alarms are mostly due to homogeneous rectangular parcels present on the ground.

By comparison with ground-truth photographs, we can enumerate three non-detected footprints on scene 1 and two on scene 2. They are due to the presence of non-rectangular buildings and to the presence of buildings with weak contrast or imperfect boundaries. If we take into account the only buildings with rectangular footprint, a good mean detection rate of 88% is obtained.

IV. PROJECTION AND REGISTRATION OF THE OPTICAL POTENTIAL FOOTPRINTS INTO SAR DATA

To combine information issued from optical and SAR images, we have to project and register homologous building primitives in a common referential.

We assign here as primitive any feature or object that characterizes the building presence in the data. In our case, during the registration process, optical primitives are the optical building edges delineating the extracted rectangular footprints and facing the SAR sensor. Their corresponding SAR primitives will be defined in Section IV-B according to the building orientation.

A two-phase process is proposed to achieve a fine registration at the metric scale.

(a) (b)

Fig. 4. Results of the building footprint extraction, superimposed to the optical image (DigitalGlobe). The extracted boundaries are represented in red. (a) Result on scene 1. (b) Result on scene 2.

A. Phase 1: Projection of the Optical Footprints Using the Optical–SAR Physical Joint Model

As described in [44], [47], and [55], the projection of points from an optical image into a SAR image needs the use of direct and inverse localization functions. These transformations are founded on the fundamental optical and SAR equations relative to the imaging systems. They permit to put in relation the 2-D pixel coordinates of a point in an image referential with its 3-D cartographic coordinates in an Earth referential (and vice versa). These functions require the knowledge of some acquisition system parameters and a height information. In our case, this height information does not need to be very accurate. It corresponds only to an approximate value that will be used to easily bound a height search interval.

In this paper, the two following models are available.

1) For the TerraSAR-X data, all the parameters of the real physical sensor model are directly known.
2) For the Quickbird data, a typical Quickbird rational polynomial coefficient (RPC) model is available; it gives a mathematical mapping between point coordinates in an Earth referential and in the optical image referential.

The height information is given by a DTM, available on the studied area and produced by the SPOT Image company.

The projection phase is realized by using an optical–SAR joint model. It allows to project the corner points at both ends of each potential footprint edge: in the first time, from the optical image into a WGS-84 Earth referential and, in the second time, from this WGS-84 Earth referential into the SAR image.

As we are dealing with satellite systems, the use of a few ground control points may be useful to improve the joint model of projection.

B. Phase 2: Fine Registration of the Optical Footprints in the SAR Image

After the phase of projection, small errors of registration are still present. They are mainly due to two reasons: First, the knowledge of the DTM heights is not accurate enough (the DTM altimetric precision is between 10 and 20 m). Second, the RPC model can induce small errors of localization in the cartographic referential. Altimetric and planimetric errors of localization in this referential cause then errors along the range and azimuth SAR axis, after the projection of corner points into SAR data.

In order to obtain a fine ground registration, a very precise ground height has to be locally estimated for the corner points. Consequently, as presented in [42], we propose to perform a phase of refinement to jointly get an accurate DTM and a good point matching.

The process uses the correspondence between the projected optical building ridges facing the sensor and their homologous SAR features.

If the building is imaged with a strictly positive aspect angle (common case), two walls (defined by two edges in the optical image) are facing the SAR sensor. The SAR primitives corresponding to these edges are thus the two bright linear SAR echoes issued from double reflections between ground and walls.

If the building is oriented parallel to the azimuth direction (particular case), only one wall is facing the SAR sensor. In this case, it provides only one double bounce stripe in the SAR image.

We assume that the studied building walls are sufficiently rough at the considered wavelength (diffuse reflection) to make both double echoes (common case) clearly visible on the SAR image, whatever the building orientation. We consider thus that even if the building walls imaged with a relatively large aspect angle tend to have a weak double stripe, the global backscattering of such building ground corners stays sufficiently important to make the double echoes still observable. In practice, this is verified on our data set. In particular, on scene 2, we can observe that, for each tall building (whose aspect angle is between 25° and 65°), both strong double bounce stripes appear.

The idea is to fit the coordinates of building corner points in the intermediate Earth referential in such a way to ensure the superposition between selected optical ridges and corresponding SAR echoes after projection into SAR data.

The registration is run through a first stage of optimization and a second stage of regularization.

1) Optimization: For each optical building footprint imaged with a strictly positive aspect angle (common case), we proceed as follows.

1) The two ridge segments $(Sab, Sac)_{\text{optical}}$ that can be imaged as ground–wall SAR echoes are designated in the optical image referential, given the SAR sensor position.
2) The ended points $(Pa, Pb, Pc)_{\text{optical}}$ of both segments $(Sab, Sac)_{\text{optical}}$ are deduced.
3) The points $(Pa, Pb, Pc)_{\text{optical}}$ are projected into the Earth referential giving the new points $(Pa, Pb, Pc)^0_{\text{Earth}}$

$$(Pa, Pb, Pc)^0_{\text{Earth}} = P_{O\to E}\left[(Pa, Pb, Pc)_{\text{optical}}\right] \quad (3)$$

where $P_{O\to E}$ is the projection operator from the optical image referential (O) into the Earth referential (E) (applied for each point with the same average height value z_{mean} issued from the DTM).

4) The optimization scheme is performed through a greedy algorithm, testing exhaustively a set of potential small displacements $(\delta_x, \delta_y, \delta_z)^k$ in the Earth referential, as described as follows:
   a) definition of the test points $(Pa, Pb, Pc)^k_{\text{Earth}}$ in the Earth referential by

$$(Pa, Pb, Pc)^k_{\text{Earth}} = T_k\left[(Pa, Pb, Pc)^0_{\text{Earth}}\right] \tag{4}$$

   where T_k corresponds to the 3-D translation operator according to the displacement $(\delta_x, \delta_y, \delta_z)^k$ in the Earth referential;
   b) obtention of their corresponding test points $(Pa, Pb, Pc)^k_{\text{SAR}}$ in the SAR image referential by

$$(Pa, Pb, Pc)^k_{\text{SAR}} = P_{E\to S}\left[(Pa, Pb, Pc)^k_{\text{Earth}}\right] \tag{5}$$

   where $P_{E\to S}$ is the projection operator from the Earth referential (E) to the SAR image referential (S) (applied for each point with the same height value $(z_{\text{mean}} + (\delta_z)^k)$);
   c) computation and storage of a registration score s^k, relative to the associated couple of segments $(Sab, Sac)^k_{\text{SAR}}$

$$s^k = \sum_{(i,j)\in(Sab\cup Sac)^k_{\text{SAR}}} \frac{r_{(i,j)}}{L_{(Sab\cup Sac)^k_{\text{SAR}}}} \tag{6}$$

   where $r_{(i,j)}$ corresponds to the radiometry of the pixel (i, j) on the SAR image and $L_{(Sab\cup Sac)^k_{\text{SAR}}}$ corresponds to the cumulated length of both segments $(Sab, Sac)^k_{\text{SAR}}$;
   d) estimation of the optimal set of displacements $(\delta_x, \delta_y, \delta_z)^*$ (ensuring the best feature matching) by score maximization

$$(\delta_x, \delta_y, \delta_z)^* = \underset{(\delta_x,\delta_y,\delta_z)^k}{\arg\max}\,(s^k) \tag{7}$$

   e) deduction of the optimal localization of $(Pa, Pb, Pc)^*_{\text{Earth}}$ in the Earth referential from (4) applied with the vector $(\delta_x, \delta_y, \delta_z)^*$;
   f) deduction of the optimal localization of their projected points $(Pa, Pb, Pc)^*_{\text{SAR}}$ in the SAR image referential from (5) applied with the optimal height $(z_{\text{mean}} + (\delta_z)^*)$.
5) The refined DTM is generated using the values $(z_{\text{mean}} + (\delta_z)^*)$ locally obtained for the footprints.
6) The set of optimal displacements $(\delta_x, \delta_y, \delta_z)^*$ in the Earth referential is converted into a set of optimal translations $(t_{\text{range}}, t_{\text{azimuth}})^*$ in the SAR image for the following step.

In the case of a building oriented parallel to the azimuth direction (particular case), the same process as before is applied but considering only the ridge segment $(Sab)_{\text{optical}}$ (and its two ended points $(Pa, Pb)_{\text{optical}}$), induced by the single wall facing the SAR sensor.

The different parameters used in this stage of optimization are the following:

1) the maximal translations $t^{\max}_{\text{range}}$ and $t^{\max}_{\text{azimuth}}$ envisaged in the SAR image referential that are fixed to a few pixels;
2) the strides s_{range} and s_{azimuth}, taken between each tested translation, that depend on the adopted compromise between registration precision and computing time.

2) Regularization: In the previous refinement, we assume that the presence of visible SAR double echoes is verified for each building. In reality, this hypothesis is not always true (in particular, in dense urban areas). We suggest thus to perform a final phase of regularization: It consists of selecting, among the optimal translations $(t_{\text{range}}, t_{\text{azimuth}})^*$ computed independently for the different footprints, the ones that appear as the most reliable and propagating them. To do this, we take into account slow variations of the DTM and locally similar errors of localization.

This stage can be summarized as follows.

In the first time, three constraints called $C_{\text{visibility}}$, $C_{\text{radiometry}}$, and C_{tendency} are introduced.

1) $C_{\text{visibility}}$ refers to the possibility, for a selected optical ridge, to be imaged as a ground–wall echo on the SAR data, given its spatial surrounding. This constraint is verified when there is, around the considered footprint, any other building close enough to hide its echo.
2) $C_{\text{radiometry}}$ refers to the optimal score of registration s^*. This constraint is verified if s^* is superior to a threshold $T_{\text{registration}}$.
3) C_{tendency} refers to the optimal translations $(t_{\text{range}}, t_{\text{azimuth}})^*$ of the considered building. This constraint is verified if the translation values $(t_{\text{range}}, t_{\text{azimuth}})^*$ are close enough to a mean tendency $(t_{\text{range}}, t_{\text{azimuth}})^*_{\text{mean}}$, provided by a local average.

In the second time, the decision about reliability is taken: When at least one constraint is not verified, the associated optimal translations are considered as not reliable. They are thus replaced by the nearest translations obtained in a neighborhood that satisfy all the constraints.

The previous refined DTM can be also corrected: Among the optimal individual heights $(z_{\text{mean}} + (\delta_z)^*)$, only those that induce reliable translations are kept. The others are replaced by using a similar strategy of propagation.

The different parameters used in this stage of regularization are the following:

1) a diameter $d_{\text{neighborhood}}$ used to define a neighborhood around each building and fixed to a few pixels;
2) the threshold $T_{\text{registration}}$ that can be learned on double echo areas on calibrated SAR data;
3) two thresholds $T^{\text{tolerance}}_{\text{range}}$ and $T^{\text{tolerance}}_{\text{azimuth}}$ used to specify the allowed gap between the values $(t_{\text{range}}, t_{\text{azimuth}})^*$ and $(t_{\text{range}}, t_{\text{azimuth}})^*_{\text{mean}}$.

C. Results on the Studied Scenes

The step of footprint projection and registration is applied on real data for both studied scenes.

TABLE II
PARAMETERS USED FOR BUILDING FOOTPRINT PROJECTION AND REGISTRATION

	Scene 1	Scene 2
$t_{range}^{max}(pixels)$	+/- 3	+/- 6
$t_{azimuth}^{max}(pixels)$	+/- 6	+/- 12
$s_{range}(pixels)$	0.25	0.25
$s_{azimuth}(pixels)$	0.25	0.25
$d_{neighbourhood}(pixels)$	5	5
$T_{registration}$	450	450
$T_{range}^{tolerance}(pixels)$	+/- 3	+/- 3
$T_{azimuth}^{tolerance}(pixels)$	+/- 6	+/- 6

The radiometric thresholds (especially $T_{registration}$) are given for Single Look Complex (SLC) SAR data coded on 32 bits.
These data are not calibrated and are directly computed from the digital number (DN) values given in amplitude (square root of the power) in TerraSAR-X data.
To obtain the associated calibrated values to radar brightness β_0 (beta nought), the power values have to be multiplied by the calibration constant k_s (here: $k_s = 1.2562 \times 10^{-5}$).
In this way, the converted values may be used for data from other sensors, which improves the portability of the algorithm.

Table II indicates the parameter values used during this step. All values are identical for both scenes, except $t_{\text{range}}^{\max}$ and $t_{\text{azimuth}}^{\max}$. Larger maximal translations have been allowed for the second scene: Indeed, as the hypothesis of locally flat ground is not verified everywhere on this scene, stronger registration errors are likely to be present after projection.

Fig. 5 presents the two maps indicating the location of the registered potential buildings in the SAR image referential.

During the registration, only translations relative to four buildings in scene 1 and to six buildings in scene 2 have been considered as reliable values. A visual analysis shows that this step has allowed a robust, fine, and global registration: After their repositioning, the optical ground segments are precisely located in the middle of the SAR double echoes (the observed registration errors are inferior to three pixels in the mean).

V. BUILDING HEIGHT ESTIMATION AND VALIDATION

This step allows to benefit from the data fusion framework. We jointly aim to confirm or not the presence of the potential buildings (phase of validation) and to estimate their elevation (phase of height retrieval).

A semi-automatic approach, using the combination of two SAR criteria, depending on the building height, is proposed.

A three-phase process is presented. In the first time, a statistical global criterion is optimized to provide a set of candidate building heights. In the second time, the introduction of a radiometric local criterion permits to select the appropriate building height and to validate or not the building presence. In the last time, a score of confidence is evaluated on the SAR image for each reconstructed building.

A. Presentation of Geometric Configurations for the Studied Buildings

The proposed methodology is based on the presence of characteristic building areas on the SAR image. We have thus to define the location and extent of these areas, given the 3-D building dimensions.

In Fig. 6, two kinds of simple geometric configurations are proposed. The first one [Fig. 6(a)] deals with the case where the building height h_{building} is lower than a certain height h_{limite}, depending on some parameters (building dimensions and sensor incidence angle [23], [36]). It is adapted for modeling low industrial buildings. The second one [Fig. 6(b)] deals with the case where the building height h_{building} is larger than h_{limite}. It is adapted for modeling high residential buildings. At the moment, only flat roofs are taken into account.

The different areas composing the building signature can be listed as follows:

1) for configuration 1: the "background area"; the "layover area" corresponding to the simultaneous backscattering from building roof, frontages, and background; the bright "double echo area"; the "single roof area" where only the roof is responding; and the "shadow area" where no object is responding;
2) for configuration 2: the "background area"; the "layover area"; the bright "double echo area"; the "frontage area" where only frontages and background are responding; and the "shadow area."

The location and the extent of each area can be calculated for a supposed building height h, from the knowledge of the building footprint location in the SAR image and for some given SAR acquisition system parameters (incidence angle and range spacing [23], [36]).

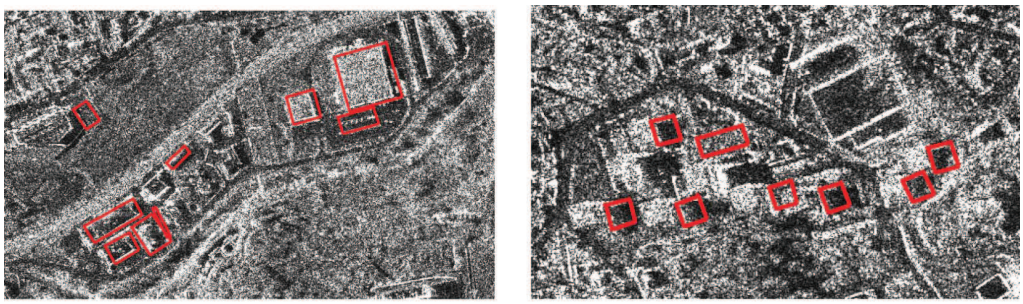
These characteristic building areas will be denoted by $CBA(h)$ in the following.

B. Phase 1: Statistical Criterion for Candidate Building Height Generation

1) Principle: The idea is to exploit the geometric extents of $CBA(h)$ to estimate the building elevation. To do so, we propose to use as statistical criterion the log-likelihood of the intensities in the SAR image for the regions defined by $CBA(h)$. For a given building and a given height h, this global log-likelihood is defined as the sum of all the log-likelihoods on the different parts in $CBA(h)$ ("background area," "layover area," etc.). For each area i in $CBA(h)$, the intensity distribution is classically modeled by a gamma density function defined by two parameters (μ_i, L_i).

The analytical expression of the log-likelihood $LL_{\text{area}}(i)$ in this area i is thus given by

$$LL_{\text{area}}(i) = N_i \left(L_i \log(L_i) - \log G(L_i)\right) - \frac{L_i}{\mu_i} \sum_{j \in i} x_j$$
$$-N_i \log(\mu_i) + (L_i - 1) \left(\sum_{j \in i} \log(x_j) - N_i \log(\mu_i) \right) \quad (8)$$



(a) (b)

Fig. 5. Results of the building footprint projection and registration, superimposed to the SAR image (Infoterra). The extracted boundaries are represented in red. (a) Result on scene 1. (b) Result on scene 2.

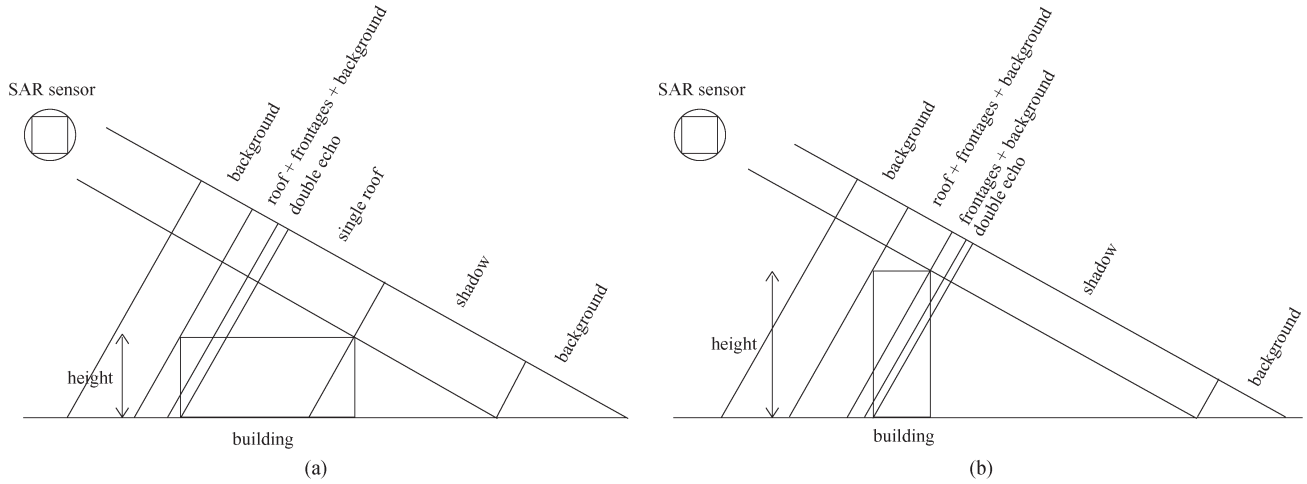


Fig. 6. Representations of the two studied configurations. The different characteristic areas composing the building signature are defined in slant range geometry on the SAR image. (a) Configuration 1: $h_{\text{building}} < h_{\text{limite}}$. (b) Configuration 2: $h_{\text{building}} > h_{\text{limite}}$.

where N_i is the number of pixels in the area i, x_j is the intensity of pixel j, and $\log G$ is the log-gamma function.

Both (μ_i, L_i) parameters are locally estimated by the maximum likelihood method.

Our final statistical criterion $LL(h)$ is defined as the opposite of the global log-likelihood

$$LL(h) = -\sum_{i \in CBA(h)} LL_{\text{area}}(i). \tag{9}$$

It will correspond to an energy to minimize with h.

2) Optimization: A global strategy of "*Height Hypothesis–Partitioning Generation–Criterion Optimization*" is employed. For each potential optical footprint, projected and registered into SAR data, the following steps are applied.

1) First, the position of the four building corners is deduced from the rectangular footprint. The building subscene, required to compute the global criterion, is defined by extending the footprint bounding box.
2) Then, the following scheme is performed:
   a) selection of a height test value h_{test} in the interval Δh of potential building heights;
   b) computation of the subscene partitioning into the characteristic areas $CBA(h_{\text{test}})$ by using a geometric SAR model of building projection;
   c) evaluation of the statistical criterion $LL(h_{\text{test}})$ on the test partitioning.
3) Finally, the curve representing $LL(h_{\text{test}})$ for all possible h_{test} values is generated. When a representative local optimum clearly appears, a strong geometric adequation exists between the predicted signature and the real one. This optimum can be local or global depending on the scene complexity.

Local optima are thus selected, giving, for the second phase, candidate building heights defined by

$$h_{\text{candidate}} = \arg\min_{h_{\text{test}}} -\text{local}\left(LL(h_{\text{test}})\right). \tag{10}$$

To improve the selection of $h_{\text{candidate}}$, two shape constraints called C_{width} and C_{height} are added on the curve. We define by $W_{\text{peak}}(h_{\text{candidate}})$ and $H_{\text{peak}}(h_{\text{candidate}})$ the width and the relative height of the optimal peak obtained for $h_{\text{candidate}}$. The constraints C_{width} and C_{height} are, respectively, satisfied if $W_{\text{peak}}(h_{\text{candidate}})$ is higher than a threshold $T^{\min}_{\text{width}}$ and if $H_{\text{peak}}(h_{\text{candidate}})$ is higher

than a threshold $T^{\min}_{\text{height}}$. This permits to delete several non-representative optima.

Lets us remark that this approach is similar to the one proposed by Brunner *et al.* [36], based on a scheme called "*hypothesis generation–rendering–matching*" for building height retrieval from SAR images. Nevertheless, some main differences can be underlined: In the approach of Brunner *et al.* [36], a matching criterion is computed to test the similarity between a simulated SAR image and the real SAR one. In our approach, no SAR data are simulated. Instead, a statistical criterion is computed directly from the pixel intensities on the real SAR image to test the adequation between a partitioning and the real building signature.

The different parameters used in this phase of candidate height generation are the following:

1) the size ϵ_{echo} used to define the bandwidth of the "double echo area" and fixed to two or three pixels;
2) the size ϵ_{box} used to extend the footprint bounding box around each building;
3) the heights $h_{\min}$ and $h_{\max}$ used to bound the interval Δh and set according to the geometric configuration;
4) a quantification step h_{step} to cover the interval Δh;
5) the thresholds $T^{\min}_{\text{width}}$ and $T^{\min}_{\text{height}}$ used to define the representativity of optimal peaks in the curve $LL(h_{\text{test}})$ and, respectively, fixed to a few meters and a few percent.

C. Phase 2: Radiometric Criterion for Final Building Height Estimation and Building Validation

1) Principle: To choose the most appropriate height among the candidate ones, we propose to combine the statistical criterion with a radiometric criterion.

The radiometric criterion tests the presence of a discontinuity between the "background area" and the "layover area." The location of this discontinuity is predicted for a given h_{test} and can be tested by computing the local contrast between both areas. This contrast is defined as the mean amplitude ratio R of pixels belonging to narrow bands on both sides of the edge. Given the *a priori* knowledge that the "background area" is darker than the "layover area," R has to be smaller than one, and, when the partitioning of the subscene matches perfectly well with the SAR image, R is minimal.

2) Combination of Both Criteria: The following scheme is performed:

1) for each $h_{\text{candidate}}$, computation and storage of the minimal ratio $R(h)$ obtained for a height h in the interval $[h_{\text{candidate}} - \delta h, h_{\text{candidate}} + \delta h]$, where δh is defined as a close neighborhood around $h_{\text{candidate}}$; this local minimum ratio, denoted by $R(h^{\text{ratio}}_{\text{candidate}})$, can be obtained for a height $h^{\text{ratio}}_{\text{candidate}}$ slightly different than $h_{\text{candidate}}$;
2) selection of $h_{\text{candidate}}$ running to the best minimal contrast $R(h^{\text{ratio}}_{\text{candidate}})$;
3) estimation of the final height $h_{\text{estimated}}$, defined as the mean between the selected $h_{\text{candidate}}$ and the associated $h^{\text{ratio}}_{\text{candidate}}$;
4) thresholding on $R(h^{\text{ratio}}_{\text{candidate}})$ to validate or not the building.

The different parameters used in this phase of height retrieval and validation are the following:

1) a size $\epsilon_{\text{discontinuity}}$ used to define the bandwidth on both sides of the discontinuity and empirically fixed to three pixels;
2) a width δh used to define a close neighborhood around $h_{\text{candidate}}$ and fixed to 2 m;
3) the threshold T_{ratio} applied on $R(h^{\text{ratio}}_{\text{candidate}})$ for validation and fixed to 0.8 by the operator to avoid too strict thresholding.

D. Phase 3: Qualification by Scores of Confidence

To provide quality measures, we propose to compute, for each reconstructed building, a score of confidence.

We define a global score of confidence S_{global}, between zero (weak confidence) and one (strong confidence), by merging individual selected scores of confidence. These individual scores, denoted by S_i (where $i = 1, \ldots, 6$), refer to the reliability ensured by the optimal peaks in the curves $LL(h_{\text{test}})$ and $R(h_{\text{test}})$.

The individual scores S_1, S_2, and S_3 take into account some aspects relative to the statistical criterion $LL(h_{\text{test}})$ as follows.

1) S_1 refers to the optimal value of the statistical criterion $LL(h_{\text{candidate}})$.
2) S_2 refers to the optimal value of $W_{\text{peak}}(h_{\text{candidate}})$.
3) S_3 refers to the optimal value of $H_{\text{peak}}(h_{\text{candidate}})$.

The individual scores S_4, S_5, and S_6 take into account some aspects relative to the radiometric criterion $R(h_{\text{test}})$ as follows.

1) S_4 refers to the optimal value of the radiometric criterion $R(h^{\text{ratio}}_{\text{candidate}})$.
2) S_5 refers to the smallest size of δh leading to a contrast smaller than T_{ratio}.
3) S_6 refers to the location gap between the selected $h^{\text{ratio}}_{\text{candidate}}$ and the height h_{test} leading to the nearest local minimum of the radiometric criterion.

S_{global} is simply defined as the mean of the scores

$$S_{\text{global}} = \frac{1}{6}\sum_{i=1}^{6} S_i. \tag{11}$$

Thresholding on S_{global} could help to detect the remaining false alarms. A manual inspection by an operator could be envisaged for buildings that have not passed the test.

E. Validation of the Methodology on Simulated Data

To validate the methodology, the approach has been first tested on two simulated SAR images of buildings.

Figs. 7(a) and (b) shows the result of building simulation according to both geometric configurations (case of a building A with a height h_{building} smaller than h_{limite} and case of a building B with a height h_{building} larger than h_{limite}). A different gamma density function has been used for each characteristic building area.

Table III presents the values of the parameters used to generate the simulated data. Buildings A and B have been simulated under a respective height hypothesis of 10 m [Fig. 7(a)] and 27 m [Fig. 7(b)].

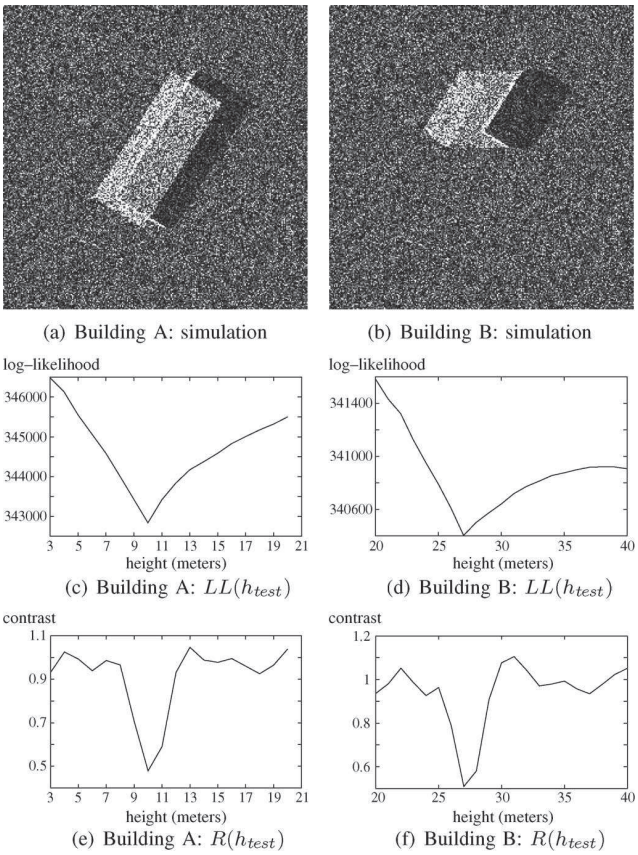


Fig. 7. Results obtained on the simulated data. (a) Building A: simulation. (b) Building B: simulation. (c) Building A: $LL(h_{\text{test}})$. (d) Building B: $LL(h_{test})$. (e) Building A: $R(h_{\text{test}})$. (f) Building B: $R(h_{\text{test}})$.

TABLE III
PARAMETERS USED FOR BUILDING SAR SIMULATION

	Building A	Building B
building height $h_{building}$	10 m	27 m
footprint width	50 pixels	25 pixels
footprint length	125 pixels	60 pixels
footprint orientation	32 degrees	32 degrees
footprint position (center)	(160, 55)	(160, 55)
h_{limite}	30 m	16 m
$(\mu_i, L_i)_{background}$	(64, 1)	(64, 1)
$(\mu_i, L_i)_{layover}$	(252, 1)	(252, 1)
$(\mu_i, L_i)_{shadow}$	(30, 1)	(30, 1)
$(\mu_i, L_i)_{double-echo}$	(500, 1)	(500, 1)
$(\mu_i, L_i)_{single-roof}$	(140, 1)	-
$(\mu_i, L_i)_{frontage}$	-	(140, 1)

The evolution curves of the energy $LL(h_{\text{test}})$ and of the contrast $R(h_{\text{test}})$ are shown in Figs. 7(c)–(f).

The proposed criteria seem quite efficient: Indeed, in both cases, a minimal value of LL and R clearly appears for a common building height and permits to retrieve the true height. Best minimal contrasts of 0.48 [Fig. 7(e)] and 0.51 [Fig. 7(f)] are obtained and allow to validate the buildings.

F. Results on the Studied Scenes

The step of building height estimation and validation is applied on real data for both studied scenes.

TABLE IV
PARAMETERS USED FOR BUILDING HEIGHT ESTIMATION AND VALIDATION

	Scene 1	Scene 2
$\epsilon_{echo}(pixels)$	3	2
$\epsilon_{box}(pixels)$	100	100
$h_{min}(meters)$	3	5
$h_{max}(meters)$	15	30
$h_{step}(meters)$	1	1
$T^{min}_{width}(meters)$	3	3
$T^{min}_{height}(percents)$	2	2
$\epsilon_{discontinuity}(pixels)$	3	3
$\delta h(meters)$	2	2
T_{ratio}	0.8	0.8

Table IV indicates the parameter values used during this step.

Table V details, for the different buildings in both scenes, the heights obtained at the different phases ($h_{\text{candidate}}$, $h^{\text{ratio}}_{\text{candidate}}$, and $h_{\text{estimated}}$). The first column refers to the building index B. The fourth column refers to the building status *Status* (validated or rejected).

The final results, presented in Fig. 8, provide, for each studied scene, the gray level map of the estimated building heights for the validated buildings. In Fig. 9, 3-D textured views of the buildings composing the scenes are proposed.

Among the 17 potential footprints on both scenes, 14 correct decisions have been taken after validation: 13 existing buildings have been correctly validated, and one non-existing building has been correctly rejected. The three wrong decisions correspond to two false alarms and one non-detection.

We can globally say that, as long as the scatter interferences between close buildings stay relatively weak, the process is able to deal with small and local intensity perturbations and succeeds in producing correct height extraction and validation results. However, when these perturbations become too important (which happens when buildings are tall and close), the process fails. For instance, this is the case of building 3 on scene 2 that has been wrongly rejected. This is due to two combined phenomena: First, the complete lack of "shadow area" (due to the superposition between the "shadow area" of the considered building and the "layover area" of the building behind) has prevented the statistical criterion from delivering accurate candidate building heights. Second, the presence of perturbations in the narrow band, used on the background side of the tested background–layover discontinuity (due to punctual bright targets and to the partial superposition between the "background area" of the considered building and the "double echo area" of another one), has prevented the radiometric criterion from estimating a correct refined height. This has led to building rejection, since the radiometric contrast is not well verified (a best minimal ratio of 0.87, higher than T_{ratio}, has been obtained).

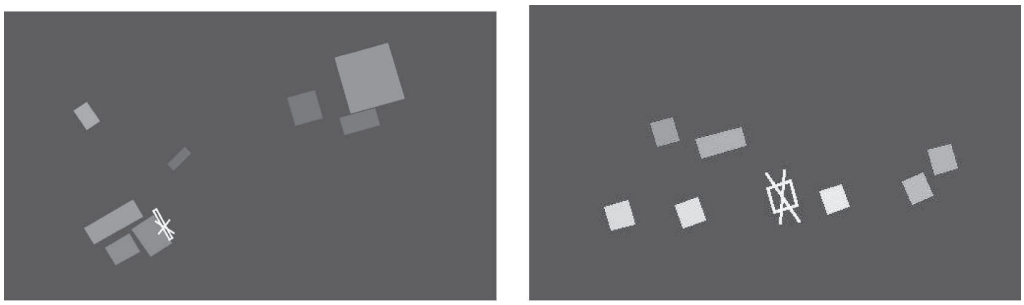
Table VI provides, for both scenes, the differences of heights between the ground-truth building heights and the final estimated building heights, for the validated buildings. The retrieved building heights appear quite relevant and precise. No general trend of over- or under-estimation is observed for the height estimation results. Root mean square errors of 1.34 and 0.85 m are obtained, respectively, on scenes 1 and 2.

The proposed 3-D reconstructed scenes are thus globally satisfying.

TABLE V
TABLES OF ESTIMATED BUILDING HEIGHTS ON SCENES 1 AND 2

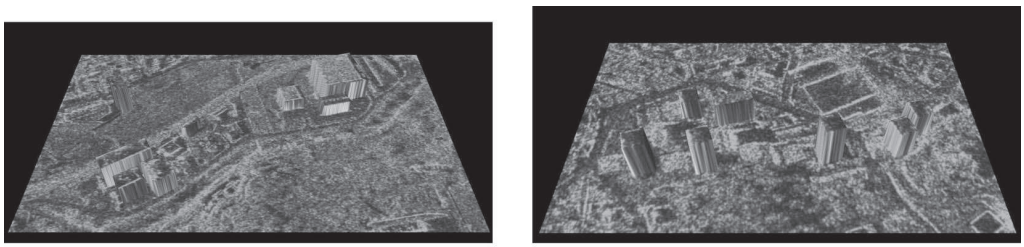
Scene 1				
B	$h_{candidate}$	$h^{ratio}_{candidate}$	$Status$ (1)	$h_{estimated}$
1	12 m	13 m	V	12.5 m
2	6 m	4 m	V	5 m
3	5 m	3 m	V	4 m
4	11 m	10 m	V	10.5 m
5	8 m	8 m	V	8 m
6	7 m	9 m	V	8 m
7	10 m	9 m	V	9.5 m
8	4 m	5 m	V	4.5 m
9	- m	- m	R	-
(1) Building status : V = validated - R = rejected				

Scene 2				
B	$h_{candidate}$	$h^{ratio}_{candidate}$	$Status$ (1)	$h_{estimated}$
1	11 m	11 m	V	11 m
2	19 m	21 m	V	20 m
3	22 m	21 m	R	21.5 m
4	21 m	21 m	V	21 m
5	15 m	13 m	V	14 m
6	22 m	23 m	V	22.5 m
7	14 m	13 m	V	13.5 m
8	15 m	15 m	V	15 m
(1) Building status : V = validated - R = rejected				



(a) (b)

Fig. 8. Results of height estimation and building validation, represented as gray level maps. A gray level scale is used to indicate the relative building heights that have been retrieved for the validated buildings. The rejected building footprints are designed by a cross. (a) Result on scene 1. (b) Result on scene 2.



(a) (b)

Fig. 9. Results of 3-D building reconstruction (Infoterra). The 3-D textured views are represented in the SAR image referential. The reconstruction has been done by combining planimetric information from the map of footprint location and altimetric information from the map of estimated heights. (a) Result on scene 1. (b) Result on scene 2.

TABLE VI
TABLES OF DIFFERENCES BETWEEN GROUND-TRUTH HEIGHTS AND ESTIMATED HEIGHTS FOR VALIDATED BUILDINGS ON SCENES 1 AND 2

Scene 1			
B	$h_{ground-truth}$ (1)	$h_{estimated}$	$Error$
1	no existing	12.5 m	-
2	4.9 m	5 m	+0.1 m
3	6.8 m	4 m	-2.8 m
4	9.1 m	10.5 m	+1.4 m
5	unknown	8 m	-
6	7.6 m	8 m	+0.4 m
7	9.3 m	9.5 m	+0.2 m
8	5.4 m	4.5 m	-0.9 m
9	no existing	(rejected)	-
Root Mean Square error of height estimation: 1.34 m.			
(1) The ground truth heights are computing by averaging the heights issued from photographs and optical shadow measures.			

Scene 2			
B	$h_{ground-truth}$ (1)	$h_{estimated}$	$Error$
1	11.2 m	11 m	-0.2 m
2	19.9 m	20 m	+0.1 m
3	19.4 m	(rejected)	-
4	20.3 m	21 m	+0.7 m
5	14.4 m	14 m	-0.4 m
6	20.6 m	22.5 m	+1.9 m
7	no existing	13.5 m	-
8	15 m	15 m	0 m
Root Mean Square error of height estimation: 0.85 m.			
(1) The ground truth heights are computing by averaging the heights issued from photographs and optical shadow measures.			

A potential improvement concerning the investigation of the radiometric criterion could be done. This criterion could be extended to other discontinuities in order to consolidate the selection of $h_{\mathrm{candidate}}$. In particular, we could focus on the edge between the "shadow area" (darker) and the "background area" (brighter).

TABLE VII
TABLE OF SCORES OF CONFIDENCE ON SCENES 1 AND 2

Scene 1		Scene 2	
B	S_{global}	B	S_{global}
1	0.66	1	0.58
2	0.66	2	0.75
3	0.5	3	-
4	0.66	4	0.83
5	0.75	5	0.58
6	0.66	6	0.75
7	0.83	7	0.5
8	0.58	8	0.75
9	-		

Finally, the associated scores of confidence S_{global} are indicated in Table VII. A short analysis can be drawn: All buildings present a score S_{global} strictly higher than 0.5 except the only two buildings (building 3 on scene 1 and building 7 on scene 2). In consequence, if a manual verification was done for all buildings presenting a score smaller than 0.5, the operator would only have to inspect these two buildings. The second one would be designated as false alarm, and the rate of correct decisions could be increased (15 correct decisions among 17).

VI. DISCUSSION ABOUT THE OPERATIONALITY OF THE COMPLETE CHAIN

A. Parameter Setting

An analysis about the stability and the sensitivity of the parameters used throughout the chain is proposed in this section.

Each of these parameters can be classified into one of the three following categories:

1) The fixed parameters: These ones are empirically set as they are very stable with different input data (whatever the sensor, the kind of studied scene, or the building configuration). This is the case particularly of the parameters used in the optimization stages (examples: $a_{\max}$, $b_{\max}$, $c_{\max}$, and $d_{\max}$ in step 1; s_{range} and s_{azimuth} in step 2; and h_{step} in step 3) or for the definition of some particular zones of interest (example: $\epsilon_{\mathrm{discontinuity}}$).
2) The *a priori* known parameters: These ones are easy to set by the operator according to the maximal dimensions and the supposed configuration of the building of interest (examples: $\epsilon_{\min}$, $\epsilon_{\max}$, $S_{\min}$, $L_{\min}$, and $W_{\min}$ in step 1 and $h_{\min}$ and $h_{\max}$ in step 3), or according to the studied scene properties such as the building density or the flat ground validity (examples: $t^{\max}_{\mathrm{range}}$, $t^{\max}_{\mathrm{azimuth}}$, $d_{\mathrm{neighborhood}}$, $T^{\mathrm{tolerance}}_{\mathrm{range}}$, and $T^{\mathrm{tolerance}}_{\mathrm{azimuth}}$ in step 2).
3) The non *a priori* known parameters: These ones are set by the operator and correspond to the most critical parameters. They are often very sensitive, and their setting has an important influence on the quality of the results. They correspond mostly to the different thresholds (examples: T_{shadows}, $T_{\mathrm{overlapping}}$, T_{Sobel}, and T_{score} in step 1; $T_{\mathrm{registration}}$ in step 2; and T_{ratio} in step 3).

The two last categories require the intervention of an operator, which justifies the "semi-automatic" adjective, attributed to the processing chain.

In the last category, two kinds of threshold values can be distinguished: first, the radiometric ones that can be locally learned for given conditions $(T_{\mathrm{shadows}}, T_{\mathrm{Sobel}})$ or that can be used for input data from different sensors, if they have been computed on calibrated data $(T_{\mathrm{registration}})$ and, second, the ones for which the setting by the operator is absolutely required ($T_{\mathrm{overlapping}}$, T_{score}, and T_{ratio}).

We have indeed realized some experiments showing that, for Quickbird data, as well as for Worldview data, the values to use for T_{shadows} and T_{Sobel} are relatively constant for different scenes inside the same image (as long as we consider the same conditions of luminance).

On the contrary, concerning the second kind of critical parameters, their sensitivity can be illustrated through the following remarks. In our studied scenes, by setting the value of T_{ratio} to 0.9 (instead of 0.8) in step 3, the wrongly rejected building on scene 2 would be validated and reconstructed, but some new false alarms could appear. In the same way, setting the values of T_{score} and $T_{\mathrm{overlapping}}$ in step 1 would have also a direct impact on the detection rate and on the false alarm rate. This shows well how important is the manual setting of such threshold values.

B. Computing Time

We propose here an evaluation of the computing time, required to perform the steps of footprint extraction, projection and registration, and height estimation and validation.

The following time values are given for a typical optical subset (256 by 256 pixels), containing a single building and issued from a manual selection.

To perform the complete processing chain of 3-D reconstruction on this subset, about 5 min is required (5 min and 13 s in CPU user time, with an AMD Opteron Processor 248, working at a frequency of 2.2 GHz).

This computing time can be decomposed as follows:

Step 1) window generation and refined footprint extraction on the optical image: about 4 min and 16 s;
Step 2) projection and registration into SAR data: about 5 s;
Step 3) height estimation and validation in the SAR image: about 52 s.

The proposed algorithms could thus be implemented to work in operational conditions.

Let us notice that the main limiting phase is the first phase of the first step (corresponding to the generation of windows of interest): Indeed, the morphological operations by reconstruction, performed with different structuring element sizes to build the DMP, require a few minutes.

VII. CONCLUSION AND FUTURE WORK

An operational processing chain, producing relevant 3-D reconstruction of buildings from HR optical and SAR imagery, has been described and illustrated on real data. It has been shown that the process of optical footprint extraction could provide precise results. The step of building validation and height retrieval, using the combination of two SAR criteria, has proved

to give reliable information. Satisfying 3-D reconstruction of buildings has been finally proposed.

The main limitation of the chain relies on the simple 3-D model adopted for building that could be complexified (introduction of more sophisticated footprint shapes such as "T-shape" or "L-shape" footprints and introduction of roofs with slopes).

Several potential applications in urban environment monitoring (such as mapping of seismic vulnerability of buildings or damage assessment after natural disaster) could benefit from the proposed method.

In futher works, a symmetrization of the complete chain could be proposed by giving an equivalent weight to the optical component and to the SAR component. This would permit to fully explore the optical–SAR complementarities. For instance, during the step of footprint extraction, we could envisage to detect windows of interest also from the SAR image. In the same way, during the step of building validation and qualification, some criteria issued from the optical image could be also considered.

Moreover, an extension of this paper could be done to deal with the case where one optical image and a couple of InSAR data are available as inputs.

REFERENCES

[1] S. Müller and D. W. Zaum, "Robust building detection in aerial images," in *Proc. Int. Arch. Photogramm., Remote Sens., Spatial Inf. Sci.*, Aug. 2005, vol. 36, pp. 143–148.

[2] A. Ferro, D. Brunner, and L. Bruzzone, "An advanced technique for building detection in VHR SAR images," in *Proc. SPIE Conf. Image Signal Process. Remote Sens.*, Sep. 2009, vol. 7477, pp. 74770V-1–74770V-12.

[3] F. Tupin, H. Maître, J. F. Mangin, J. M. Nicolas, and E. Pechersky, "Detection of linear features in SAR images: Application to road network extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 2, pp. 434–453, Mar. 1998.

[4] J. D. Wegner and U. Soergel, "Bridge height estimation from combined high-resolution optical and SAR imagery," in *Proc. ISPRS Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, 2008, vol. 37, pp. 1071–1076.

[5] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, pt. 1, no. 9, pp. 1940–1949, Sep. 2003.

[6] J. Chanussot, J. A. Benediktsson, and M. Fauvel, "Classification of remote sensing images from urban areas using a fuzzy possibilistic model," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 40–44, Jan. 2006.

[7] C. Baillard and H. Maître, "3D reconstruction of urban scenes from aerial stereo imagery: A focusing strategy," *Comput. Vis. Image Understand.*, vol. 76, no. 3, pp. 244–258, Dec. 1999.

[8] C. Tison, F. Tupin, and H. Maître, "A fusion scheme for joint retrieval of urban height map and classification from high-resolution interferometric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 496–505, Feb. 2007.

[9] R. Bolter, "Reconstruction of man-made objects from high resolution SAR images," in *Proc. IEEE Aerosp. Conf.*, 2000, vol. 3, pp. 287–292.

[10] C. A. Lin and R. Nevatia, "Building detection and description from a single intensity image," *Comput. Vis. Image Understand.*, vol. 72, no. 2, pp. 101–121, Nov. 1998.

[11] K. Karantzalos and N. Paragios, "Recognition-driven two-dimensional competing priors toward automatic and accurate building detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 133–144, Jan. 2009.

[12] B. Sirmaçek and C. Ünsalan, "A probabilistic framework to detect buildings in aerial and satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 211–221, Jan. 2011.

[13] N. Chehata, F. Jung, M. Pierrot-Deseilligny, and G. Stamon, "A region-based matching approach for 3D-roof reconstruction from high-resolution satellite stereo pairs," in *Proc. DICTA*, Dec. 2003, vol. 2, pp. 889–898.

[14] A. K. Shackelford, C. H. Davis, and X. Wang, "Automated 2-D building footprint extraction from high-resolution satellite multi-spectral imagery," in *Proc. IEEE IGARSS*, 2004, vol. 3, pp. 1996–1999.

[15] Y. Wei, Z. Zhao, and J. Song, "Urban building extraction from high-resolution satellite panchromatic image using clustering and edge detection," in *Proc. IEEE IGARSS*, 2004, vol. 3, pp. 2008–2010.

[16] J. Xiaoying and H. Curt, "Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information," *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 14, pp. 2196–2206, 2005.

[17] S. Lefèvre and J. Weber, "Automatic building extraction in VHR images using advanced morphological operators," in *Proc. URBAN*, Apr. 2007, pp. 1–5.

[18] Y. Wang and H. Liu, "Semi automatic extraction of building information and variation detection from high resolution remote sensing images," in *Proc. SPIE Conf.*, vol. 6419, *Geoinformat.: Remotely Sensed Data Inform.*, 2006, pp. 64190T-1–64190T-11.

[19] N. Chehata, M. P. Desseilligny, F. Jung, and G. Stamon, "Extraction of 3D primitives from stereopairs of satellite images for automatic reconstruction of buildings," in *Proc. IAPR Conf. Workshop Mach. Vis. Appl.*, Dec. 2002, pp. 636–639.

[20] M. Ortner, X. Descombes, and J. Zerubia, "Building outline extraction from digital elevation models using marked point processes," *Int. J. Comput. Vis.*, vol. 72, no. 2, pp. 107–132, Apr. 2007.

[21] D. Flammanc, G. Maillet, and H. Jibrini, "3D city models: An operational approach using aerial images and cadastral maps," in *Proc. ISPRS Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, Sep. 2003, vol. 34, pp. 53–58.

[22] F. Taillandier, "Automatic building reconstruction from cadastral maps and aerial images," in *Proc. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, 2005, vol. 36, pp. 105–110.

[23] A. J. Bennett and D. Blacknell, "The extraction of building dimensions from high resolution SAR imagery," in *Proc. Int. Radar Conf.*, 2003, pp. 182–187.

[24] M. Quartulli and M. Datcu, "Stochastic geometrical modeling for built-up area understanding from a single SAR intensity image with meter resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 9, pp. 1996–2003, Sep. 2004.

[25] T. Balz and N. Haala, "SAR-based 3D-reconstruction of complex urban environments," in *Proc. IAPRS Workshop 3-D Reconstruct. From Airborne Laser Scanner InSAR Data*, 2003, vol. 34, pp. 181–185.

[26] E. Simonetto, H. Oriot, and R. Garello, "Rectangular building extraction from stereoscopic airborne radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2386–2395, Oct. 2005.

[27] P. Gamba, B. Houshmand, and M. Saccani, "Detection and extraction of buildings from interferometric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 611–617, Jan. 2000.

[28] S. Guillaso, L. Ferro-Famil, A. Reigber, and E. Pottier, "Building characterization using L-band polarimetric interferometric SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 3, pp. 347–351, Jul. 2005.

[29] A. Thiele, E. Cadario, K. Schulz, U. Thoennessen, and U. Soergel, "Building recognition from multi-aspect high-resolution InSAR data in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, pt. 1, no. 11, pp. 3583–3593, Nov. 2007.

[30] F. Xu and Y. Q. Jin, "Automatic reconstruction of building objects from multi-aspect meter-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, pt. 2, no. 7, pp. 2336–2353, Jul. 2007.

[31] F. Tupin, "Extraction of 3D information using overlay detection on SAR images," in *Proc. IEEE GRSS/ISPRS Joint Workshop Data Fusion Remote Sens. Over Urban Areas. URBAN*, 2003, pp. 72–76.

[32] V. Lemoigne and F. Tupin, "Statistical polygonal snakes for 3D building reconstruction using high resolution SAR data," in *Proc. Urban Remote Sensing Joint Event. URBAN*, Apr. 2007, pp. 1–5.

[33] G. Franceschetti, A. Iodice, and D. Riccio, "A canonical problem in electromagnetic backscattering from buildings," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1787–1801, Aug. 2002.

[34] G. Franceschetti, A. Iodice, D. Riccio, and G. Ruello, "SAR raw signal simulation for urban structures," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, pt. 1, no. 9, pp. 1986–1995, Sep. 2003.

[35] R. Guida, A. Iodice, and D. Riccio, "Height retrieval of isolated buildings from single high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2967–2979, Jul. 2010.

[36] D. Brunner, G. Lemoine, L. Bruzzone, and H. Greidanus, "Building height retrieval from VHR SAR imagery based on an iterative simulation and matching technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1487–1504, Mar. 2010.

[37] H. Sportouche, F. Tupin, and L. Denise, "Building detection by fusion of optical and SAR features in metric resolution data," in *Proc. IEEE IGARSS*, Jul. 2009, vol. 4, pp. IV-769–IV-772.
[38] F. Cellier, H. Oriot, and J. M. Nicolas, "Hypothesis management for building reconstruction from high resolution InSAR imagery," in *Proc. IEEE IGARSS*, 2006, pp. 3639–3642.
[39] U. Stilla, U. Soergel, and U. Thoennessen, "Potential and limits of InSAR data for building reconstruction in built-up areas," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 1/2, pp. 113–123, Jun. 2003.
[40] F. Tupin and M. Roux, "3D information extraction by structural matching of SAR and optical features," in *Proc. ISPRS Int. Arch. Photogramm. Remote Sens.*, 2004, vol. 35, pp. 489–494.
[41] Z. Junjie, D. Chibiao, Y. Hongjian, and X. Minghong, "3D reconstruction of building based on high-resolution SAR and optical images," in *Proc. IEEE IGARSS*, 2006, pp. 3794–3797.
[42] H. Sportouche, F. Tupin, and L. Denise, "Building extraction and 3D reconstruction in urban areas from high-resolution optical and SAR imagery," in *Proc. Urban Remote Sens. Joint Event URBAN*, May 2009, pp. 1–11.
[43] H. Sportouche and F. Tupin, "A processing chain for simple 3D reconstruction of buildings in urban scenes from high resolution optical and SAR images," presented at the 8th European Conference on Synthetic Aperture Radar(EUSAR), Aachen, Germany, Jun. 2010.
[44] F. Tupin and M. Roux, "Markov random field on region adjacency graph for the fusion of SAR and optical data in radargrammetric applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1920–1928, Aug. 2005.
[45] F. Tupin, "Merging of SAR and optical features for 3D reconstruction in a radargrammetric framework," in *Proc. IEEE IGARSS*, 2004, vol. 1, pp. 89–92.
[46] J. D. Wegner, U. Soergel, and A. Thiele, "Building extraction in urban scenes from high resolution InSAR data and optical imagery," in *Proc. Urban Remote Sens. Joint Event. URBAN*, May 2009, pp. 1–6.
[47] F. Tupin and M. Roux, "Detection of building outlines based on the fusion of SAR and optical features," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 1/2, pp. 71–82, Jun. 2003.
[48] V. Poulain, J. Inglada, M. Spigai, J. Y. Tourneret, and P. Marthon, "Fusion of high resolution optical and SAR images with vector databases for change detection," in *Proc. IEEE IGARSS*, Jul. 2009, vol. 4, pp. IV-956–IV-959.
[49] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, May 2010.
[50] J. Inglada and A. Giros, "On the possibility of automatic multi-sensor image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2104–2120, Oct. 2004.
[51] T. D. Hong and R. A. Schowengerdt, "A robust technique for precise registration of radar and optical satellite images," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 5, pp. 585–593, 2005.
[52] T. Balz, "SAR simulation based change detection with high-resolution SAR images in urban environments," in *Proc. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, 2004, vol. 35, pp. 472–477.
[53] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.
[54] M. Dalla Mura, J. A. Benediktsson, L. Bruzzone, and F. Bovolo, "A novel technique based on morphological filters for change detection in optical remote sensing images," in *Proc. AIP Conf. 6th Int. WIO*, 2007, vol. 949, pp. 75–82.
[55] I. Tannous and B. Pikeroen, "Parametric modeling of spaceborne SAR image geometry. Application: SEASAT/SPOT image registration," *Photogramm. Eng. Remote Sens.*, vol. 60, no. 6, pp. 755–766, 1994.

Hélène Sportouche (M'11) received the Engineer and the Research Master degrees in optics and image processing from the École Centrale Marseille, Marseille, France, in 2007 and the Ph.D. degree in signal and image processing from Telecom ParisTech, Paris, France, in 2010.

She is currently a Postdoctoral Researcher with the Département Traitement du Signal et des Images, Telecom ParisTech, Paris, France. Her research interests are image analysis and understanding for remote sensing applications, in particular, feature extraction, object detection, 3-D reconstruction, and multi-sensor registration, on high-resolution optical and SAR data in urban areas.

Florence Tupin (SM'07) received the Engineering and Ph.D. degrees in signal and image processing from Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1994 and 1997, respectively.

Since November 1998, she has been with the Department Traitement du Signal et des Images, Telecom ParisTech, Paris, where she was an Associate Professor and has been a Professor of signal processing and computer vision since 2009. From 1997 to 1998, she was with SAGEM in fingerprint recognition. Her research interests are image analysis and interpretation, 3-D reconstruction, Markov random fields, and synthetic aperture radar, particularly for urban remote sensing applications. From 2005 to 2007, she was an Associate Editor of the French journal *Signal Processing*.

Dr. Tupin was the Chair of the Urban Remote Sensing Joint Event held in Paris in 2007. Since 2003, she has been also a member of the technical committees "Pattern Recognition in Remote Sensing" of the International Association for Pattern Recognition and "URBAN," the biennial GRSS/ISPRS Joint Workshops on "Remote Sensing and Data Fusion over Urban Areas" from 2004 to 2006. She currently serves as an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Léonard Denise (M'09) received the Engineer degree from the École Nationale Supérieure d'Ingénieurs de Constructions Aéronautiques, Toulouse, France, in 1984.

He has worked on several remote sensing projects, conducting studies in the areas of SAR system analysis, simulation, and processing for civilian and military applications. He is currently a Technical Expert with the Département IMINT, Thales Communications, Massy, France, and is involved in the development of multi-sensor image exploitation systems, more particularly for advanced applications, including SAR interferometry and radargrammetry, photogrammetry, 3-D reconstruction, recognition, and change detection.