

# Building Extraction and 3D Reconstruction in Urban Areas from High-Resolution Optical and SAR Imagery

(Invited Paper)

Helene SPORTOUCHE<sup>\*†</sup>, Florence TUPIN<sup>\*</sup> and Leonard DENISE<sup>†</sup>

<sup>\*</sup> Institut TELECOM; TELECOM ParisTech; CNRS LTCI; 75013, Paris, France

Telephone: +33 1 45 81 72 45, Mobile : +33 6 89 78 17 05, Fax : +33 1 45 81 37 94

Email: helene.sportouche@telecom-paristech.fr

Email: florence.tupin@telecom-paristech.fr

<sup>†</sup> THALES Communications; Land and Joint Systems; 91883, Massy, France

Telephone: +33 1 69 75 36 27, Fax: +33 1 69 75 31 65

Email: leonard.denise@fr.thalesgroup.com

**Abstract**— In this paper, we propose to investigate the joint use of high-resolution optical and SAR data, for building extraction and 3D reconstruction in large urban areas. A sequence of methods providing, in a semi-automatic way, the building detection and reconstruction is presented. Potential building footprints are first extracted on an optical image by a two-phases process (coarse detection and boundaries refinement). The framework of fusion with SAR data is then developed. A first step of registration allows us to get a fine superposition of optical and SAR building features. Then, we show how to take benefit from the introduction of SAR data: The proposed methodology, based on the optimization of criteria referring to building SAR characteristics, allows us to simultaneously deal with the building presence validation and with the height retrieval.

## I. INTRODUCTION

The new generation of satellite optical and SAR sensors (TerraSAR-X, Quickbird, CosmoSkyMed ...), delivering images with metric or sub-metric resolution, offers great challenges in the field of building extraction and 3D building reconstruction in urban areas.

Some well known technics have already been proposed in the fields of optical or SAR imagery to deal with this problematic and have proved to provide efficient results. In the optical field, approaches of accurate building reconstruction have been developed in the frameworks of stereoscopy [1] and monoscopy with additional input data [2]. In the SAR field, some process for building recognition and height estimation have been proposed in interferometry [3] [4] [5] [6] [7], in radargrammetry [8], in the case of a single input data [9] [10] [11] [12] [13] and also with multi-aspects SAR data [14] [15].

Nevertheless, two main difficulties can be pointed out. First, optical images are not always available, depending on weather conditions. Secondly, SAR interferometry or radargrammetry still give limited results (noisy and incomplete data). In such a context, some technics for urban structures (building, bridges...) detection and height retrieval using the

complementarity of both optical and SAR data have been explored [16] [17]. Structure shapes and locations are often provided by optical data while height information is given by a couple of SAR data [18] [19] [20] [21].

In this paper, we propose to investigate the joint use of only one HR optical image and one HR SAR image for 3D building reconstruction. The objective is to propose a sequence of methods providing, in a semi-automatic way, the extraction and the reconstruction of buildings in large urban areas. We see how it is possible to extract and combine two kinds of information from both data. We assume that our 3D building model is a simple parallelepiped (vertical frontages and flat roof), described by two kinds of parameters: its planimetric rectangular boundary and its height.

The proposed approach is decomposed into the following steps: In section II, we describe a process based on morphological and geometrical tools in order to detect and precisely extract bidimensionnal building boundaries from monoscopic optical images. The obtained map, identifying potential rectangular building location, is used as a starting point for the following. Then, the framework of fusion with SAR data is developed in order to validate building detection and to get height information. A first step of optical footprints projection into the SAR geometry is performed. The methodology used for the projection and the registration of optical / SAR features is thus described in section III. The projection step requires a ground height information provided by a DTM (Digital Terrain Model) and the knowledge of sensor acquisition system parameters. The registration step enables us to get a finer superposition of optical and SAR building ground features. In section IV, we show how to take benefit from the introduction of the SAR data. The proposed method, based on the optimization of criteria referring to building characteristics, allows us to simultaneously deal with the bidimensionnal building "classification" aspect (building

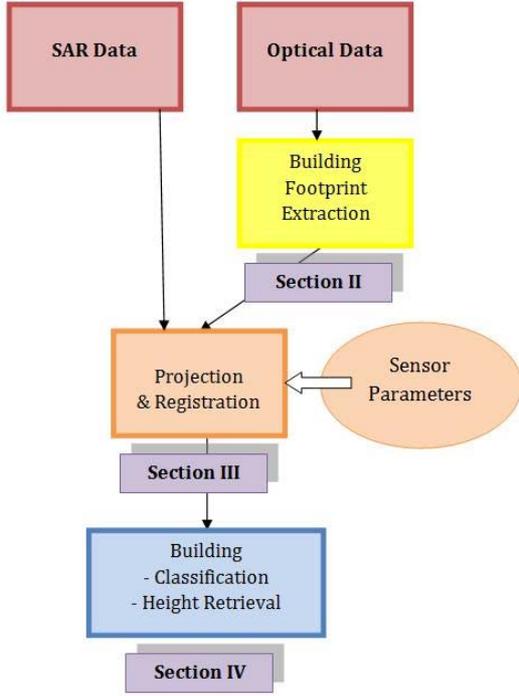


Fig. 1. Complete Process Flow

validation or rejection) and with the building height retrieval. A chart flow of the whole process is represented in Fig. 1. Results of this approach are presented on an optical Quickbird image and on a radar TerraSAR-X image, located in the urban area of Marseille in France.

## II. BUILDING BOUNDARIES EXTRACTION FROM MONOSCOPIC OPTICAL DATA

The proposed methodology aims to detect and extract quasi rectangular building boundaries in an optical urban scene. As we are mostly dealing with images whose look angle is nearly vertical (nadir), the planimetric location of the extracted boundaries, in the  $(x,y)$  coordinates system, corresponds to building roof borders as well as building ground footprints. We propose a process in two main phases, providing successively a coarse map identifying potential building presence (section II-A) and a refined map referring precisely to building boundaries (section II-B).

### A. Phase1: Global Building Detection

A combination of morphological and geometrical tools, such as Differential Morphological Profile (defined in II-A.1) and Overlapping Coefficient (defined in II-A.2), is used to perform the building detection phase. A process flow of the method is presented in II-A.2.

1) *Creation of the Differential Morphological Profile on the optical image:* The concept of Differential Morphological Profile (DMP), introduced in [22], has already been used for some applications in urban areas such as classification

[23] [24] and change detection [25]. The DMP, providing information about the size and the contrast of an object, is built by using opening and closing operators by reconstruction (i.e. with a geodesic metric) in order to preserve structure shapes. A multi-scale approach with different Structuring Elements (SE) is adopted in order to explore a large range of potential object sizes.

Lets us define the following derivative vectors  $DMP_\gamma(x, y)$ ,  $DMP_\phi(x, y)$  and  $DMP(x, y)$ , respectively called Derivative of the Opening Profile, Derivative of the Closing Profile and the Differential Morphological Profile DMP (sometimes also called Derivative of the Morphological Profile), defined at the pixel  $(x, y)$  of the image  $I$  by:

$$DMP_\gamma^i(x, y) = MP_\gamma^i(x, y) - MP_\gamma^{i-1}(x, y) \quad (1)$$

with

$$MP_\gamma^i(x, y) = \gamma_{i*s} [I(x, y)] \quad , \forall i \in [1, p]$$

$$DMP_\phi^i(x, y) = MP_\phi^i(x, y) - MP_\phi^{i-1}(x, y) \quad (2)$$

with

$$MP_\phi^i(x, y) = \phi_{i*s} [I(x, y)] \quad , \forall i \in [1, p]$$

and

$$DMP^i(x, y) \hat{=} \begin{cases} DMP_\phi^i(x, y) \quad , \forall i \in [1, p] \\ DMP_\gamma^i(x, y) \quad , \forall i \in [-p, -1] \end{cases} \quad (3)$$

where

- $\gamma$  is the opening operator by reconstruction,
- $\phi$  is the closing operator by reconstruction,
- $i$  is the vector index (iteration),
- $(2p + 1)$  is the DMP dimension,
- $s$  is the radius of a disc SE and  $(i * s)$  is the SE size used at the iteration  $i$ . In the following, we will call  $\epsilon$  the SE size ( $\epsilon = i * s$ ).

When a SE reaches in the DMP the characteristic size of an object, all its belonging pixels receive the same 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 structures brighter or darker than surrounding. Objects are thus progressively removed at different levels of the DMP and a stack of simplified images is created.

In the presented process, we take advantage of the properties of image simplification and shape preservation, allowing us to introduce a geometrical criterion on the a priori building shape. We assume indeed that our buildings can be modeled on the optical image by homogeneous rectangular shapes likely to be clearly identified at different levels of the DMP.

TABLE II

EVALUATION OF THE BUILDING DETECTION PROCESS (PHASE 1)

Detection rate	False alarm rate	Averaged misclassified pixel rate
Study for the optical extract number 1 (47 buildings)		
75 %	31 %	25 %

2) *Hierarchical process in two passes, based on a geometrical adequation test*: The hierarchical process can be decomposed into the four following parts:

- Pre-Processing: The shadows, which are clearly visible on the optical data, are extracted by a simple thresholding on the grey level value and are then masked in order to make the building identification easier.
- Pass 1: The following processing is iteratively applied to the different images composing the DMP, from images corresponding to the larger SE to the ones corresponding to the smaller SE. For each test area, present in the considered image of the DMP and corresponding to a non zero grey level local value, a potential rectangular building bounding box is built using axis of inertia of the test area. An overlapping coefficient, corresponding to the geometrical adequation between the rectangular reference area and the test area, and defined as the ratio of surfaces between both of them, is computed. If the coefficient is higher than a threshold, the generated area is validated and masked for the next iteration.
- Pass 2: This pass is applied for building with imperfect underlying boundaries. In order to clearly draw the limit between an imperfect area and its surrounding, classical erosion and dilation with small SE are performed on each image of the DMP. Then, the same processing as the one of Pass 1 is performed on this modified DMP.
- Post-Processing: A simple thresholding on the minimal size of the validated buildings is realized in order to reduce the number of false alarms.

A summary of the different parameters used to perform this phase of detection is presented in table I. For each detected building, a score of reliability can be deduced from the chosen value of the overlapping coefficient  $C_{overlapping}$ . The parameters  $\epsilon_{min}$ ,  $\epsilon_{max}$  and  $s$ , controlling the extremal sizes and sampling of SE, can be easily adjusted in a priori known intervals of values at a given resolution.

3) *Results and evaluation of the detection on Quickbird images*: The results of this first process phase, applied to 3 urban subscenes, are displayed in Fig. 2. In the following, we will respectively call these subscenes: optical extract number 1, 2 and 3.

A good global detection is obtained even if this one is more difficult for buildings with weak contrast or imperfect boundaries. Of course, the main limitation is the a priori rectangular shape. False alarms are mostly due to homogeneous rectangular parcels present on the ground. Since elevation information is not available (one single optical image with a nadir look angle), these false alarms can not be distinguished from real elevated building roofs. An other effect can be pointed out: in the case of roofs composed with different homogeneous rectangular subparts (variation of illumination according to roof slope, striped texture...), several similar juxtaposed rectangles are detected. A quantitative qualification of this first phase is proposed in table II for the extract number 1, that presents an important diversity of buildings in terms of size, contrast and texture. The detection rate and the false

alarm rate (given in percents) are computed in relation to the whole number of buildings. The misclassified pixel rate (given in percents) is computed in terms of number of pixels. The detection rate is quite satisfying. The important averaged rate of misclassified pixels (computed on the pixels belonging to the correctly detected buildings) is mostly due to defaults of the orientation or dimensions of the detected rectangles.

### B. Phase 2: Refinement of Rectangle Position

The geometrical information of the rectangles (orientation, dimensions) provided by the previous detection phase, is not always accurate enough. To improve this, a second step is performed.

1) *Local refinement by Hough Transform*: The previous obtained coarse map is used as an input for this boundaries extraction phase and allows us to directly focus on likely building areas. Whereas a region based approach had been proposed in the first phase, a complementary contour based approach is here adopted. This approach aims to obtain a fine superposition between final extracted boundaries and radiometric pixel discontinuities. The method is globally based on the generation of rectangles, using a local Hough Transform with constraints and a radiometric criterion. The algorithm is applied around each rectangle provided by the previous phase of detection and can be divided into three steps:

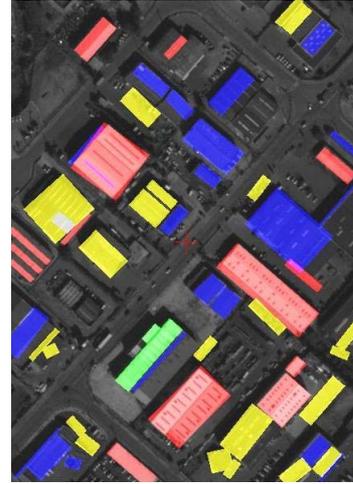
- Computation of the local Hough accumulator  $H^w(\rho, \theta)$  in a research windows  $w$ , by using the binary gradient image (after thresholding) and where each couple  $(\rho, \theta)$  defines a unique potential linear segment in a polar coordinates system. The used research window  $w$  is defined around each potential rectangle issued from the phase 1. As the windows are spatially limited, this step is quite fast.
- Selection of lines of interest from the local accumulator according to geometrical constraints: At each iteration, some lines are selected if their corresponding value in the accumulator is higher than the current threshold  $T^w$  value. The threshold  $T^w$  is decreased, while the selected lines do not fill some geometrical conditions. The iterative

TABLE I  
PARAMETERS USED FOR THE BUILDING DETECTION PROCESS (PHASE 1)

<i>Parameter</i>	<i>Description</i>	<i>Standard Value</i>	<i>Meaning</i>
<i>Coverlapping</i>	Overlapping Coefficient	from 0.7 to 1	Defining the reliability of the shape adequation
$\epsilon_{min}$	SE minimal size in DMP	from 5 to 15 m	Relative to the minimal building size
$\epsilon_{max}$	SE maximal size in DMP	from 15 to 100 m	Relative to the maximal building size
<i>s</i>	Radius of the elementary disc SE	from 1 to 5 m	Relative to the SE sampling (SE size = $i*s$ at iteration $i$ )
$\epsilon_{dilate}$	SE size for the dilation operator	from 1 to 3 m	Helping to the imperfect boundaries definition
$\epsilon_{erode}$	SE size for the erosion operator	from 1 to 3 m	Helping to the imperfect boundaries definition
$T_{shadows}$	Threshold for the shadow detection	30 - 40 (255 grey levels)	Allowing to build a mask of shadows
$S_{minimal}$	Minimal building size	from 5 to 20 m	Helping to reduce false alarm rate (small rectangular parcels)



(a) extract 1 ©DigitalGlobe



(b) result of the detection process



(c) extract 2



(d) result of the detection process



(e) extract 3 ©DigitalGlobe



(f) result of the detection process

Fig. 2. Detection Process (Phase 1). Results on the extract 1 (Quickbird sensor, resolution: 0.6 m, Marignane - France), on the extract 2 (airborne sensor, submetric resolution, St-Mandrier - France) and on the extract 3 (Quickbird sensor, resolution: 0.6 m, Marseille - France). The results issued from pass 1 of the process are respectively represented in red and blue for the opening and closing parts of the DMP. The ones from pass 2 are in green and yellow.

process is done according to the following scheme:

while

$$(N \leq N_{max} \text{ and } K \leq K_{max} \text{ and } T^w \geq T_{min}^w)$$

do

$$\left\{ \begin{array}{l} T^w = T^w - 1 \\ N = N + 1 \text{ if a new line is selected} \\ K = K + 1 \text{ if a new orthogonal angle is detected} \end{array} \right\}$$

where  $N$  is the number of selected lines and  $K$  the number of orthogonal angles potentially existing between two selected lines.

- Generation of rectangles from these selected lines by maximization of a criterion  $C^{gen}$ : Among all the selected lines, the ones associated to orthogonal angles are kept. For each detected orthogonal angle, a set of potential rectangles are generated by adding prolongations (width and length) of different sizes. Finally, a  $C^{gen}$  criterion is introduced to find the optimal combination of sizes.  $C^{gen}$  is defined as the number of pixels of the binary gradient image belonging to the generated rectangle. The rectangle maximizing the score  $C^{gen}$  is conserved.

During this local extraction process, several cases can occur according to the geometrical configuration existing between the selected lines:

- No orthogonal angle  $K$  has been detected: It means that there are no lines, whose directions are orthogonal, among all the selected lines. No orthogonal angle is available as input for the rectangle generation process. No rectangle is found.
- A single or several orthogonal angles  $K$  have been detected: one or a set of optimal rectangles are saved at the end of the optimization step.

#### 2) Results of the position refinement in Quickbird images:

The result of this second phase, applied to the extract number 3, is displayed in Fig. 3. Table III represents the number of optimal rectangles attributed at the end of the phase 2 to each initial research window resulting from phase 1. The values taken for the parameters  $N_{max}$ ,  $K_{max}$  and  $T_{min}^w$  were in this case respectively 5, 2 and 10. At the end of phase 1, 12 potential rectangles have been detected and have consequently provided 12 research windows used as input of phase 2. At the end of phase 2, a whole set of 11 optimal labeled rectangles have been obtained. The 3 research windows that have induced no optimal rectangle in phase 2 correspond to false alarms from phase 1. (The orthogonality constraint introduced in phase 2 is not satisfied in these rejected research windows). The 9 others research windows have made possible the extraction of optical potential rectangular buildings and have provided accurate information about their boundaries (width, length, orientation and center position of the rectangles). The coordinates of the four building ground corners (or

TABLE III  
REFINEMENT PROCESS (PHASE 2)

Research windows $w$	Orthogonal angles	Generated rectangles
Study for the optical extract number 3		
$w1$	2	Label A
$w2$	2	Label B
$w3$	2	Label C
$w4$	3	Labels D,E,F
$w5$	2	Label G
$w6$	2	Label H
$w7$	0	-
$w8$	2	Label I
$w9$	2	Label J
$w10$	0	-
$w11$	2	Label K
$w12$	0	-

roof corners), available on this refined map, will be used as a starting point for the fusion with SAR data.

### III. PROJECTION AND REGISTRATION STEPS FOR THE FUSION WITH SAR DATA

#### A. Projection of the Optical Building Footprints into the SAR Image

1) *Model of projection based on the fundamental SAR equations and on the optical acquisition system:* As described in [17] and [19], the projection of points from an optical image into a SAR image needs the use of direct and inverse transformation functions. They are based on the fundamental SAR equations and on the optical acquisition system and allow us to compute the 3D coordinates of an intermediate point in a common cartographic referential (geolocation). The projection requires the knowledge of the sensor acquisition system parameters and a height information. In this part of the work, this height information does not need to be very accurate. It corresponds only to an approximate height value, that will be used to bound the height search interval during the features registration step.

In the present study, the two following models are available:

- For the TerraSAR-X data, all the parameters of the real physical sensor model are known and can be used in the transformation functions.
- For the Quickbird data, a typical Quickbird RPC (Rational Polynomial Coefficients) Model, providing a mathematical mapping from object space coordinates to image coordinates, is available instead of the QuickBird Rigorous Sensor Model. It allows us to directly perform, under a height hypothesis, the projection step from the optical referential into the 3D referential.

The height information is provided by a SPOT DIMAP DEM.

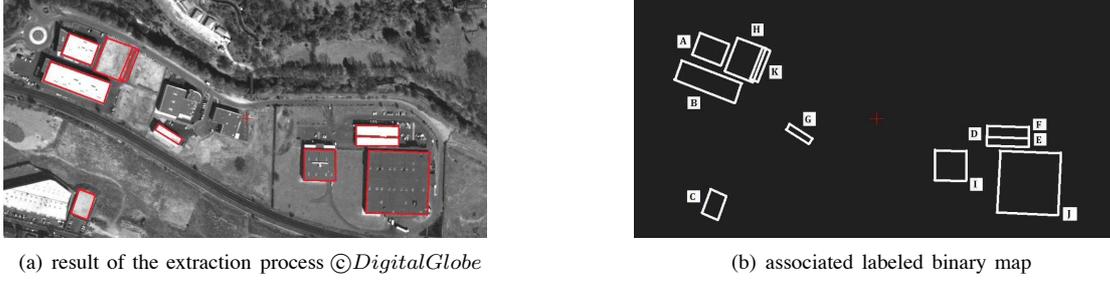


Fig. 3. Refinement Process (Phase 2). Result on the extract 3. The refined boundaries are represented in red (Fig. 3(a)) or in white (Fig. 3(b)).

2) *Improvement of the joint optical / SAR model by the introduction of a few ground control points:* As we are dealing with satellite systems, the use of ground control points may be useful to improve the projection step. The introduction by an operator of a few control points allows us to improve the estimation of the joint optical / SAR model.

3) *Results on Quickbird and TerraSAR-X images:* The building ground boundaries, previously extracted in the optical image, are projected into the SAR data. The result of this projection, applied to the optical extract number 3 and projected into the SAR image 1 (located in the same urban area), is displayed in Fig. 5(a) and 5(b).

### B. Refinement of the Registration

In spite of the previous step of joint model refinement by control points, small errors of registration are still present. They are mainly due to two reasons: First, we do not have a sufficiently accurate knowledge of the DTM. The altimetric precision of the DIMAP DEM is approximatively of 10-20 meters and the given height values can take into account the elevation, in addition to the ground height. Secondly, the optical RPC model used for the projection of optical rectangles into the cartographic referential can induce small errors of geolocation in the intermediate cartographic referential. Concerning the errors due to the DTM, an altimetric difference  $\delta z$  in the cartographic referential induces an error along the range axis during the projection step into the SAR image. Concerning those due to the optical model of projection, a planimetric difference  $(\delta x, \delta y)$  in the cartographic referential induces errors along the range and azimuth axis during the projection step into the SAR image. As we are going to project features supposed to be on ground, a very fine local ground height estimation is required. Consequently, a step of registration refinement is performed in order to jointly get a refined DTM and a good matching of optical and SAR ground features.

Two kinds of data are used as input for the registration:

- The list of the bidimensionnal coordinates of each rectangle in the optical image referential (vector data).
- The SAR image (pixel data).

Under the assumption that double bounce linear echoes from ground / wall building corners are mainly present in the SAR image, our registration process is based on the fine

superposition of the projected optical building ground ridges with these very bright SAR echoes. As the initial optical - SAR registration resulting from the improved model of projection is already good, only local small displacements  $\{\delta x, \delta y, \delta z\}$  are envisaged into the cartographic referential. These  $\{\delta x, \delta y, \delta z\}$  displacements are translated in the SAR image referential by small displacements along the x range axis and the y azimuth axis. The research of optimal translations in the SAR image referential is run by a first step of energy maximization on the SAR data, followed by a second step of regularization.

1) *Optimization:* The principle is the following one:

First, some relevant segments, likely to appear as ground / wall bright echoes on the SAR image, are selected. Indeed, as the SAR sensor position is known, it is possible to predict, for each optical rectangle, the two segments (or the three associated terminal points) that can be imaged on the SAR data as ground ridges segments (or respectively as ground corner points) with a ground elevation. These selected ground segments are going to be used to refine the local ground altitude.

Secondly, the following optimization scheme represented in Fig. 4 is applied for each optical rectangle, in order to perform the DTM refinement (z coordinate) and the planimetric geolocation refinement (x and y coordinates) in the cartographic referential. For each building  $B_k$ :

- The three selected "ground" points  $(P1, P2, P3)_{optic}$  are projected into the 3D cartographic referential, using the height  $z_{mean}$  given by the averaged height value computed for the three points. The obtained points are called  $(P1, P2, P3)_{carto}$ .
- Then, the optimization step on the  $\{\delta x, \delta y, \delta z\}$  displacements values used for the projection of  $(P1, P2, P3)_{carto}$  into the SAR image is applied. For each  $\{\delta x, \delta y, \delta z\}$  test values, the points  $(P1', P2', P3')_{carto} = (P_i')_{carto}$  are defined by their coordinates in the cartographic referential:

$$\begin{aligned} x_i'_{carto} &= x_i_{carto} + \delta x, \forall i = 1, 2, 3 \\ y_i'_{carto} &= y_i_{carto} + \delta y, \forall i = 1, 2, 3 \\ z_i'_{carto} &= z_{mean} + \delta z, \forall i = 1, 2, 3 \end{aligned}$$

These new points  $(P_i')_{carto}$  are projected into the SAR image referential and the obtained points are called  $(P1, P2, P3)_{SAR}$ . The idea consists to optimize the displacements values, until that the two selected ground

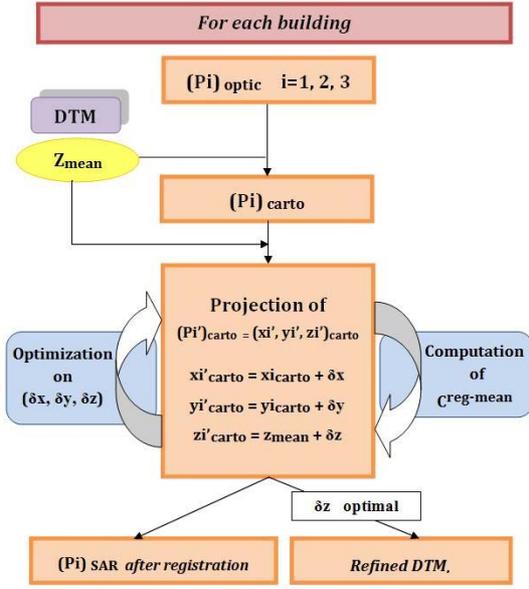


Fig. 4. Registration Process Flow

SAR segments defined by  $(P1, P2, P3)_{SAR}$  are precisely superposed with the bright ground / wall linear echoes, present on the SAR image. To do this, a criterion of registration  $C^{reg-mean}$  is introduced and computed on  $(P1, P2, P3)_{SAR}$ . It corresponds to the radiometric mean on the SAR image of all the pixels belonging to the two selected ground SAR segments.

- The optimal values  $\{\delta x, \delta y, \delta z\}$  maximizing  $C^{reg-mean}$  are saved and called  $\{\delta x, \delta y, \delta z\}^{B_k}_{optimal}$ . These optimal cartographic displacements are translated in the SAR image referential by equivalent translations  $(t_x, t_y)^{B_k}_{optimal}$  along the range axis and the azimuth axis. The obtained height values  $(z_{mean} + \delta z)^{B_k}_{optimal}$  allow us to build a locally refined DTM at the building scale.

2) *Regularization Step*: The previous DTM refinement and planimetric geolocation refinement are done for each building separately. In order to avoid DTM and geolocation errors, a regularization step is introduced taking into account slow variations of the DTM and locally similar errors of geolocation for the different buildings. When these assumptions are violated, it probably means that the registered segments of a given building have been attracted by bright structures different from the ground / wall echo of the considered building, such as single roof echoes. Consequently, a regularization step has to be performed in order to propagate in a neighbourhood the most reliable translation values. The following proposed scheme is applied to achieve, as well as possible, a robust registration under constraints.

Let us introduce the three following conditions  $L^{energy}$ ,  $L^{tendency}$  and  $L^{visibility}$ . These constraints refer essentially to two aspects: on one hand, the probability to image on the SAR data the ground wall echo issued from a building  $B_k$  according to its surrounding ( $L^{visibility}$ ) and, on the other hand, the

TABLE IV  
CRITERIA USED FOR A ROBUST BUILDING REGISTRATION

$B_k$	$C^{reg-mean}$	$L^{energy}$	$L^{tend.}$	$L^{visib.}$	$(t_x, t_y)$
Study for the optical extract number 3 associated to the SAR image 1					
A	409	0	-	-	not used
B	2618	1	1	1	used
C	193	0	-	-	not used
D	175	0	-	-	not used
E	292	0	-	-	not used
F	161	0	-	-	not used
G	891	1	0	-	not used
H	374	0	-	-	not used
I	980	1	1	1	used
J	833	1	1	1	used
K	306	0	-	-	not used

confidence that can be granted to the optimal translations values  $(t_x, t_y)^{B_k}_{optimal}$  issued from the previous registration step for the building  $B_k$  ( $L^{energy}$ ,  $L^{tendency}$ ). These conditions are defined by:

- $L^{energy}(B_k)$  is equal to one if  $C^{reg-mean}(B_k)$  is superior to  $C^{reg-mean}_{threshold}$  and zero else.
- $L^{tendency}(B_k)$  is equal to one if  $(t_x, t_y)^{B_k}_{optimal}$  is close to the mean tendency of the optimal translations obtained for the neighbour buildings and zero else.
- $L^{visibility}(B_k)$  is equal to one if there is no building very close to  $B_k$  and potentially able to hide its ground / wall echo and zero else.

In the case of one building  $B_k$ , for which the value of  $L^{energy}(B_k)$  or  $L^{tendency}(B_k)$  or  $L^{visibility}(B_k)$  is equal to zero, we consider that the associated optimal translation value  $(t_x, t_y)^{B_k}_{optimal}$  is not robust enough. Thus, we propose to replace it by the mean value of the optimal translation values obtained for the neighbour buildings, which satisfy all the conditions.

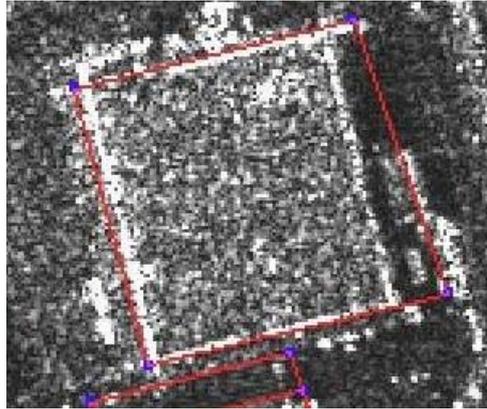
3) *Results on Quickbird and TerraSAR-X images*: The result of this registration process, applied to the extract number 3, is displayed for one building in Fig. 5(d) (zoom on the building J). The table IV represents, for the different buildings present on the extract number 3, the scores obtained for the criterion  $C^{reg-mean}(B_k)$  and for the tested conditions  $L^{energy}(B_k)$ ,  $L^{tendency}(B_k)$  and  $L^{visibility}(B_k)$ . At the end of the registration under constraints, only the translation values corresponding to the buildings labeled B, I and J have been taken into account as robust values for the propagation registration strategy. The comparison of the results obtained before and after the registration (Fig. 5(c) and 5(d)) shows that this step has made possible a very fine repositioning of the optical ground segments. They are now precisely located in the middle of the SAR linear double bounce echoes.



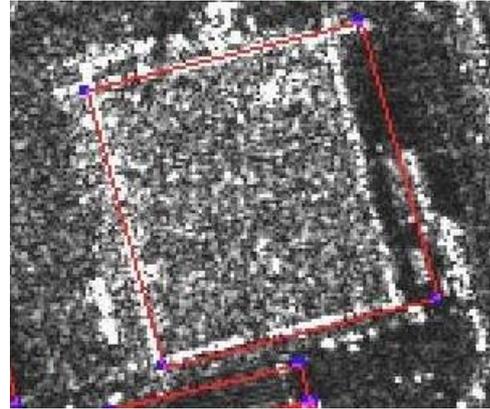
(a) projection with the original joint optical / SAR model



(b) projection with the improved model (with 5 control points)



(c) zoom on building J after projection with the improved model



(d) zoom on building J after registration with regularization

Fig. 5. Projection and Registration Process. Result on the SAR image 1 (TerraSAR-X sensor ©Infoterra, resolution: 1.1 m, area of Marseille - France) and zoom on the building J. The SAR image 1 corresponds to the same urban scene as the optical extract 3. The projected and registered rectangular building boundaries resulting from the projection (Fig. 5(a), 5(b) and 5(c)) and from the registration process (Fig. 5(d)) are represented in red. The building corner points are in blue.

#### IV. BUILDING CLASSIFICATION AND HEIGHT RETRIEVAL IN THE FRAMEWORK OF OPTICAL - SAR DATA FUSION

The framework of fusion with SAR data is now developed in order to benefit from information provided by both kinds of data. The SAR classification aims to reduce the false alarm rate, by validating or rejecting buildings previously detected on the optical data. This step is based on the presence of characteristic building features on the SAR image and enables us to deal also with building height retrieval.

##### A. Definition of the Characteristic Building Areas and Validation Criteria

At the end of the registration step, optical building footprints have been correctly projected and located on the SAR image. When these potential footprints refer to true buildings, a good superposition is obtained between the bright ground / wall echo and the corresponding optical footprint.

In the proposed validation process, adapted SAR criteria are used to verify the presence of building characteristic areas around these potential footprints. Since building features, such as single roof textures, border roof echoes or shadows, can appear very differently according to many considerations

(building roof material, sensor incidence angle, sensor wavelength...), different scenarii can happen. As represented in Fig. 6, in this paper, two kinds of simple configurations are considered:

- 1) Model 1 (Eg.: tiled or textured roofs): The characteristic building features clearly visible on the SAR image are the bright roof texture area, the shadow area, the double bounce echo area coming from the ground / wall corner and the background area.
- 2) Model 2 (Eg.: flat roofs): The characteristic building features clearly visible on the SAR image are the simple bounce echo area on the roof, the double bounce echo area coming from the ground / wall corner and the background area.

As the SAR sensor incidence angle and the geometrical building parameters (dimensions, orientation) are known, it is possible to predict the location of each characteristic building area, in function of the building height. This building height corresponds to the proper building elevation existing in addition to the ground altitude provided by the refined DTM.

For each configuration, an adapted criterion, function of the test building height  $h$ , is proposed and used as evidence of

the model adequation. These criteria are based on the proper mean radiometry characterizing each area.

- 1) For the model 1:  $C_1(h)$  refers to the radiometric discontinuity between the "roof texture area" and the "shadow area". It is defined as the mean contrast between the two areas, computed on narrow pixel bands on both sides of the roof limit.
- 2) For the model 2:  $C_2(h)$  refers to the radiometric discontinuities between the "background area" and the "simple bounce echo area on roof" and between the "simple bounce echo area on roof" and the "not visible roof texture area". It is linked to the mean contrasts between the two first areas or between the two second areas. It is computed on narrow pixel bands on both sides of the left or right limit associated to the "simple bounce echo area".

### B. Building / Non Building Classification Process

The complete methodology used in the classification and height estimation process is described in Fig 7. For each building  $B_k$ :

- A height test value  $h$  belonging to an  $[h_{min}, h_{max}]$  interval is selected.
- The relative roof and shadow areas are generated on the SAR image from the registered building footprints.
- Both criteria  $C_1(h)$  and  $C_2(h)$  are computed on these generated test areas.
- The optimal criteria values ( $C_1$  and  $C_2$ ) and height values ( $h_{C_1}$  and  $h_{C_2}$ ) maximizing them are saved.
- The final decision of building validation is then taken. A thresholding step on the saved criteria values is applied in order to guarantee a relevant radiometric modelization: If at least one of both criteria runs to an optimal value higher than a chosen threshold, the building  $B_k$  presence is validated. Else, we conclude to a false alarm.

### C. Height Estimation

From the previous optimization step, building elevations can be directly deduced from the optimal height test values  $h_{C_1}$  and  $h_{C_2}$ . For each validated building, at least one of both criteria makes a potential height value available. Sometimes, it can happen that the criteria propose two very different heights for a same building  $B_k$ . In this case, the intervention of an operator will be required, in the frame of a semi supervised building reconstruction process. Indeed, this one will have to finally choose among both models the one that visually seems the more appropriate to the  $B_k$  building features present in the SAR image.

### D. Results on Quickbird and TerraSAR-X Images

The result of the classification and height retrieval process, applied to the optical extract number 3 and to the SAR image 1, is presented in table V, where the following information are available for each building: optimal value of  $C_1$ ,  $C_2$ ,  $h_{C_1}$  and  $h_{C_2}$  at the end of the optimization step, building status (validated / rejected) at the end of the classification, and final

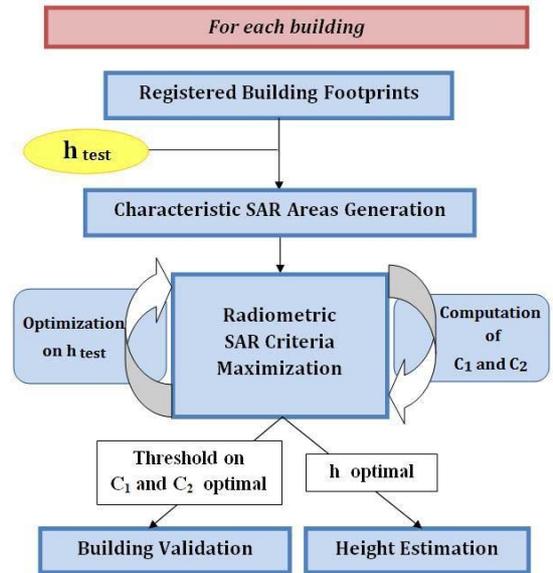


Fig. 7. Building Classification and Height Retrieval Process Flow

TABLE V  
BUILDING VALIDATION AND HEIGHT ESTIMATION

Optimal values of $C_i$ and $h_{C_i}$ , $\forall i = 1, 2$						
$B_k$	$C_1$	$h_{C_1}$	$C_2$	$h_{C_2}$	status	$h_{final}$
Study for the optical extract number 3 associated to the SAR image 1						
A	10.6	6 m	5.6	12 m	validated	12 m
B	4.3	4 m	3.5	15 m	validated	15 m
C	$\leq 4$	-	$\leq 3$	-	rejected	-
D	4.2	6 m	$\leq 3$	-	validated	6 m
E	$\leq 4$	-	$\leq 3$	-	rejected	-
F	5.4	5 m	$\leq 3$	-	validated	5 m
G	$\leq 4$	-	6.3	8 m	validated	8 m
H	$\leq 4$	-	$\leq 3$	-	rejected	-
I	5.5	10 m	$\leq 3$	-	validated	10 m
J	7.6	14 m	3	15 m	validated	14-15 m
K	5.1	12 m	$\leq 3$	-	validated	12 m

estimated building height at the end of the retrieval process. An example of building roof and shadow areas generation is displayed with the optimal height value in Fig. 8(a) and 8(b) (zoom on the building J).

### E. Analysis

The results of validation and height retrieval (table V) are globally satisfying: Among the 11 present buildings, three of them (labeled C, E and H) are rejected while they indeed correspond to false alarms issued from the optical detection process. The buildings C and H are rectangular parcels, whereas the building E corresponds only to a subpart of the whole building D. A building height estimation has been achieved for the eight validated buildings. For two of them (labeled A and B), the selection of the adapted model by an operator has been

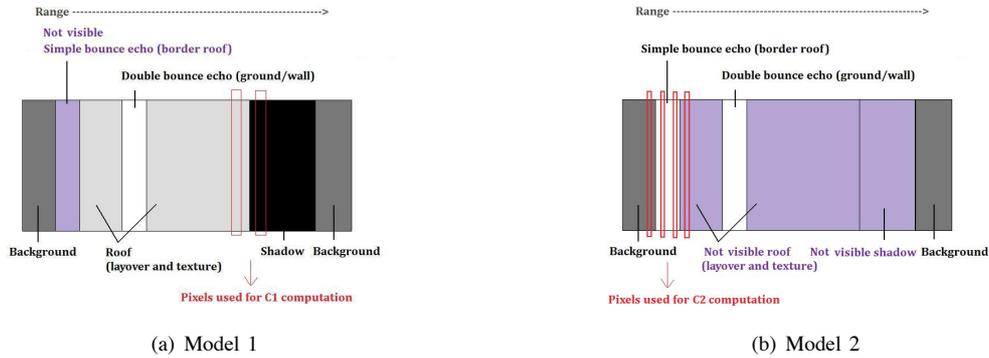
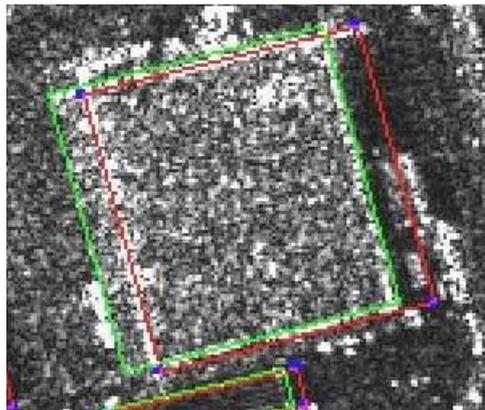
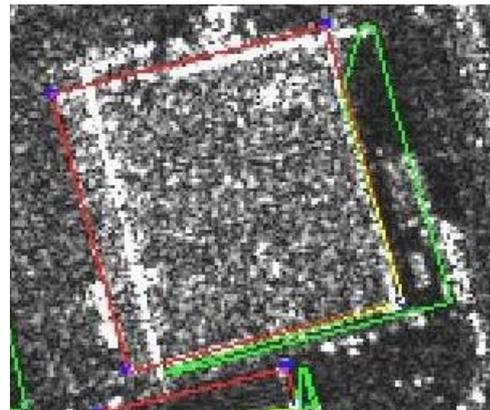


Fig. 6. Characteristic building areas modelization.



(a) Zoom on building J. Result of ground boundaries (in red) and roofs (in green) generation



(b) Zoom on building J. Result of roofs (in red) and shadows (in green) generation

Fig. 8. Characteristic areas generation. Result on the SAR image 1 ©Infoterra: Zoom on building J. The characteristic areas (ground boundaries, roofs and shadows) generated for the validated buildings with the optimal building heights are represented in red or green on the SAR image. The building corner points are in blue (Fig. 8(a) and 8(b)).

required.

## V. CONCLUSION AND FURTHER WORKS

In the frame of optical and SAR data fusion, we have proposed an approach for 3D building reconstruction in large urban areas. This framework has allowed us to benefit from the complementarities of two kinds of sensors and has provided, in operational conditions, good results, which could not be achieved by the only use of an optical or a SAR image. A methodology for planimetric building extraction from monoscopic optical images has been presented in a first part. Then, the frame of fusion has been developed: We have shown how to concretely manage with the different building features present in both data. The steps of projection and features registration have been first presented. Then, we have seen that the introduction of SAR data could be used for the building presence validation and could add relevant height information for the reconstruction. The satisfying obtained results have been illustrated on a Quickbird and a TerraSAR-X images, located in a same urban area.

In our future works, the optical - SAR fusion framework will be further developed. First, a general approach based on

the optimization of a global criterion (such as the Likelihood) will be envisaged. Then, a more elaborate interpretation of elevated building structures could be done by the introduction of complex models taking into account all the elements likely to appear on the SAR image (chimneys, eaves, sophisticated roof shapes...). Their detection could be done by specific tools on SAR data. Their matching with optical elements could help to increase the detection rate and could make the height retrieval more robust. Moreover, in the case of optical and SAR data acquired at different time, the presence of bright structures present on one single image would allow us to deal with the problem of change detection.

## REFERENCES

- [1] N. Chehata, F. Jung, M. Deseilligny, and G. Stamon, "A region-based matching approach for 3D-roof reconstruction from high-resolution satellite stereo pairs," in *DICTA*, 2003, pp. 889–898.
- [2] F. Lafarge, P. Trontin, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny, "An automatic building extraction method: Application to 3D-city modeling," 2006.
- [3] C. Tison, F. Tupin, and H. Maitre, "Retrieval of building shapes from shadows in high resolution SAR interferometric images," in *2004 IEEE International Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings*, vol. 3, 2004.

- [4] F. Cellier, H. Oriot, and J. Nicolas, "Hypothesis management for building reconstruction from high resolution InSAR imagery," in *IEEE International Conference on Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006*, 2006, pp. 3639–3642.
- [5] U. Stilla, U. Soergel, and U. Thoennessen, "Potential and limits of InSAR data for building reconstruction in built-up areas," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 1-2, pp. 113–123, 2003.
- [6] P. Gamba, B. Houshmand, and M. Saccani, "Detection and extraction of buildings from interferometric SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 1, pp. 611–617, 2000.
- [7] S. Guillaso, L. Ferro-Famil, A. Reigber, and E. Pottier, "Building characterization using L-band polarimetric interferometric SAR data," *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 3, pp. 347–351, 2005.
- [8] E. Simonetto, H. Oriot, and R. Garelo, "Rectangular building extraction from stereoscopic airborne radar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 10, pp. 2386–2395, 2005.
- [9] F. Tupin, "Extraction of 3D information using overlay detection on SAR images," in *Remote Sensing and Data Fusion over Urban Areas, 2003. 2nd GRSS/ISPRS Joint Workshop on*, 2003, pp. 72–76.
- [10] G. Franceschetti, A. Iodice, D. Riccio, and G. Ruello, "SAR raw signal simulation for urban structures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9 Part 1, pp. 1986–1995, 2003.
- [11] G. Franceschetti, R. Guida, A. Iodice, D. Riccio, G. Ruello, and U. Stilla, "Building feature extraction via a deterministic approach: application to real high resolution SAR images," in *IEEE International Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007*, 2007, pp. 2681–2684.
- [12] D. Brunner, G. Lemoine, and L. Bruzzone, "Extraction of building heights from VHR SAR imagery using an iterative simulation and match procedure."
- [13] A. Bennett, D. Blacknell, and Q. Ltd, "The extraction of building dimensions from high resolution SAR imagery," in *Radar Conference, 2003. Proceedings of the International*, 2003, pp. 182–187.
- [14] A. Thiele, E. Cadario, K. Schulz, U. Thonnessen, and U. Soergel, "Building recognition from multi-aspect high-resolution insar data in urban areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 11 Part 1, pp. 3583–3593, 2007.
- [15] E. Dai, Y. Jin, T. Hamasaki, and M. Sato, "Three-Dimensional Stereo Reconstruction of Buildings Using Polarimetric SAR Images Acquired in Opposite Directions," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 236–240, 2008.
- [16] U. Soergel, A. Thiele, H. Gross, and U. Thoennessen, "Extraction of Bridge Features from high-resolution InSAR Data and optical Images," *Urban Remote Sensing Joint Event, 2007*, pp. 1–6, 2007.
- [17] F. Tupin and M. Roux, "Detection of building outlines based on the fusion of SAR and optical features," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 1-2, pp. 71–82, 2003.
- [18] 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 Transactions on Geoscience and Remote Sensing*, vol. 45, no. 2, p. 496, 2007.
- [19] F. Tupin and M. Roux, "Markov random field on region adjacency graph for the fusion of SAR and optical data in radargrammetric applications," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 8, pp. 1920–1928, 2005.
- [20] F. Tupin, "Merging of SAR and optical features for 3D reconstruction in a radargrammetric framework," in *2004 IEEE International Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings*, vol. 1, 2004.
- [21] Z. Junjie, D. Chibiao, Y. Hongjian, and X. Minghong, "3D Reconstruction of Building Based on high-resolution SAR and Optical Images," in *IEEE International Conference on Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006*, 2006, pp. 3794–3797.
- [22] M. Pesaresi and J. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 309–320, 2001.
- [23] J. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9 Part 1, pp. 1940–1949, 2003.
- [24] J. Chanussot, J. Benediktsson, and M. Fauvel, "Classification of remote sensing images from urban areas using a fuzzy possibilistic model," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 40–44, 2006.
- [25] M. Dalla Mura, J. Benediktsson, L. Bruzzone, and F. Bovolo, "A Novel Technique Based on Morphological Filters for Change Detection in Optical Remote Sensing Images," in *AIP Conference Proceedings*, vol. 949, 2007, p. 75.