A fusion scheme for joint retrieval of urban map and classification from high resolution interferometric SAR images

C. Tison, F. Tupin, H. Maître

A Fusion Scheme for Joint Retrieval of Urban Height Map and Classification From High-Resolution Interferometric SAR Images

Céline Tison, Florence Tupin, and Henri Maître

***Abstract*—The retrieval of 3-D surface models of the Earth is a major issue of remote sensing. Some nice results have already been obtained at medium resolution with optical and radar imaging sensors. For instance, missions such as the Shuttle Radar Topagraphy Mission (SRTM) or the SPOT HRS have provided accurate digital terrain models. The computation of a digital surface model (DSM) over urban areas is the new challenging issue. Since the recent improvements in radar image resolution, synthetic aperture radar (SAR) interferometry, which had already proved its efficiency at low resolution, has provided an accurate tool for urban 3-D monitoring. However, the complexity of urban areas and high-resolution SAR images prevents the straightforward computation of an accurate DSM. In this paper, an original high-level processing chain is proposed to solve this problem, and some results on real data are discussed. The processing chain includes three main steps, namely: 1) information extraction; 2) fusion; and 3) correction. Our main contribution addresses the merging step, where we aim at retrieving both a classification and a DSM while imposing minimal constraint on the building shapes. The joint derivation of height and class enables the introduction of more contextual information. As a consequence, more flexibility toward scene architecture is possible. First, the initial images (interferogram, amplitude, and coherence images) are converted into higher-level information mapping with different approaches (filtering, object recognition, or global classification). Second, these new images are merged into a Markovian framework to jointly retrieve an improved classification and a height map. Third, DSM and classification are improved by computing layover and shadow from the estimated DSM. Comparison between shadow/layover and classification allows some corrections. This paper mainly addresses the second step, while the two others are briefly explained and referred to already published papers. The results obtained on real images are compared to ground truth and indicate a very good accuracy in spite of limited image resolution. The major limit of DSM computation remains the initial spatial and altimetric resolutions that need to be made more precise.**

***Index Terms*—Classification, height map, Markovian fusion, synthetic aperture radar (SAR) interferometry, urban areas.**

I. Introduction

THE EXTRACTION of 3-D town models is a major issue for many applications, such as, for example, environment or urban planning. Thanks to the recent improvement of synthetic aperture radar (SAR) image resolution, SAR interferometry can now address this issue. Taking also into account the launching of future SAR missions (SAR Lupe, CosmoSkymed, TerraSAR-X), the evaluation of the potential of interferometry over urban areas is a subject of major concern. This paper presents an original and flexible method to extract a digital surface model (DSM) from a high-resolution interferogram over urban areas. Due to their complexities, a dedicated scheme is required.

This paper is deliberately restricted to the use of one single interferometric data take per scene to fully assess the potential of interferometry. The challenge has also been the development of a method with no restriction on building shapes. Actually, urban architecture is so diverse that minimal hypotheses on building shape are required.

A. Interferometry and the Urban Area Context

An interferogram is the phase difference of two SAR images that are acquired over the same scene with slightly different incidence angles. Under certain coherence constraints, this phase difference (the interferometric phase) is linked to scene topography [1], [2]. The interferometric phase ϕ and the corresponding coherence ρ are, respectively, the phase and the magnitude of the normalized complex hermitian product of two initial SAR images (s_1 and s_2). To reduce noise, an averaging over an $L \times L$ window is added as

$$\rho e^{j\phi} = \frac{\sum_{i=1}^{L^2} s_1(i) s_2^*(i)}{\sqrt{\sum_{i=1}^{L^2} |s_1(i)|^2 \sum_{i=1}^{L^2} |s_2(i)|^2}}. \tag{1}$$

ϕ has two contributions, namely: 1) the orbital phase ϕ_{orb}, which is linked to natural variations of the line-of-sight vector, and 2) the topographical phase ϕ_{topo}. By Taylor expanding to first order, the height h of every pixel is proportional to ϕ_{topo} and depends on the wavelength λ, the sensor target distance R, the perpendicular baseline $B_\perp$, and the incidence angle θ, i.e.,

$$h = \frac{\lambda}{2\pi p} \frac{R \sin\theta}{B_\perp} \phi_{\text{topo}} \tag{2}$$

with p equal to 2 for the monostatic case and to 1 for the bistatic case.

ϕ_{orb} is only geometry dependent and can be easily removed from ϕ [2]. Therefore, in the following, the interferometric

Manuscript received November 10, 2005; revised September 29, 2006.

C. Tison is with the Centre National d'Etudes Spatiales, DCT/SI/AR, 31401 Toulouse Cedex 4, France (e-mail: celine.tison@cnes.fr).

F. Tupin and H. Maître are with GET, Ecole Nationale Superieure des Telecommunications, CNRS UMR 5141, 75013 Paris, France.

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.2006.887006

phase should be understood as the topographic phase (the orbital phase was removed previously). The height is derived from this phase. In the fusion scheme, the height map is used rather than the interferometric phase.

Although (2) looks simple, direct inversion does not lead to an accurate DSM. The first reason is knowledge of the phase modulo 2π, which requires a phase unwrapping step. The height corresponding to a phase equal to 2π is called the ambiguity altitude. This ambiguity altitude is much higher than the heights of buildings, which prevents phase wrapping. For this reason, the topic of phase unwrapping is not addressed in this paper.

For high-resolution images of urban areas, the difficulties arise from geometrical distortions (layover, shadow), multiple reflections, scene geometry complexity, and noise. As a consequence, high-level algorithms are required to overcome these problems and to have a good understanding of the scene. Height filtering and edge preservation require specific processing for the different objects of the scene (e.g., a building with a roof should not be filtered the same way as vegetation). The challenge is to get both an accurate height and an accurate shape description of each object in the scene.

B. State of the Art

High-resolution SAR images remain quite new for the scientific community, and as a consequence, only a small number of teams have access to such data. Therefore, literature on DSM computation from SAR interferometry is only at its beginning. So far, four kinds of methods have been proposed.

1) Shape from shadow [3]. Building outlines are estimated from the shadows detected in the amplitude image, whereas the interferogram provides an average height for each footprint. At least two (ideally four) amplitude images are required with optimal view angles to detect all the building edges. This requirement is very demanding and is not very realistic when working with spaceborne data takes.
2) Roof filtering [4]. The interferogram provides a noisy height map that is filtered out by looking for horizontal planes (roof buildings); these planes are initialized by 3-D segments. These first two methods are only efficient for large and isolated buildings.
3) Stochastic geometry [5]. The position and the shape of each building are optimized by a function linking the amplitude, coherence, and interferogram. Because of computational time constraints, the building shape model is restricted to a unique one. As a consequence, a strong *a priori* assumption is made in favor of urban architecture.
4) Global scene reconstruction based on a primary classification [6], [7]. This approach is the most flexible and operational (at least [7] reference). It links 3-D reconstruction and classification. At a first step, no assumption on the building shapes are required. Nevertheless, the significant results obtained by Soergel *et al.* [7] rely on the merging of several interferograms over the same scene and on a rectangular shape model for buildings.

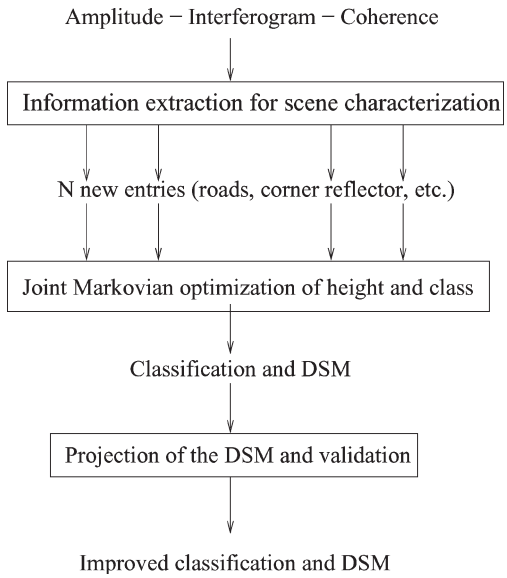


Fig. 1. Global scheme of the proposed method for DSM estimation over urban areas. The height estimation is processed jointly with a classification, as these two pieces of information are deeply linked.

In this paper, the input data over a scene are deliberately limited to an interferometric couple, and no constraint on building shapes is considered. Actually, in an operational context, the user has to work with only one interferogram per scene. In addition, town architecture is highly diverse: it cannot be restricted to one building model. This framework led us to select the fourth approach. However, instead of dealing with the classification and the DSM estimation separately, a joint computation of height and class is proposed. In fact, class and height have strong interactions that should be taken into account to improve the global scene recovery.

C. Proposed Method

Global processing is divided into three main steps (Fig. 1). Since the original SAR data are difficult to interpret, new inputs are preliminarily derived from pattern recognition methods, denoising, classification, etc., to get higher-level information. This step, which is briefly described in Section III, mainly refers to previous work. In addition, the algorithms proposed for this step and the results should be considered only as open options. Users are free to develop their own tools to derive first-step information with no impact on the global processing chain architecture.

As a second step, all these new images are merged into a Markovian framework to jointly provide a classification and a height map. The merging method is inspired by Tupin *et al.* [8]. First, an oversegmentation of the scene is applied to define regions on which classification and height recovery will be applied. This region partitioning allows for the reduction of computation time and for the description of region interactions. The joint optimization of height and class is defined in a Markovian framework using the new entries (obtained from the first step) as the observation field (Section IV). The global architecture of this second step is completely independent of the number and content of the inputs. Therefore, the result

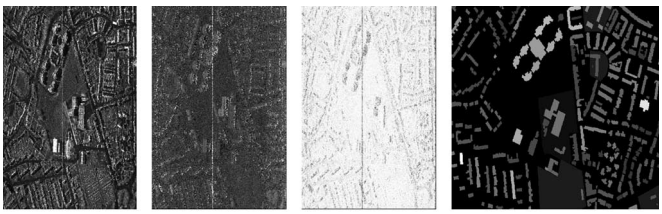


Fig. 2. Presentation of the available data takes (Bayard district). (a) Multilook amplitude. (b) Interferogram. (c) Coherence. (d) Ground truth (IGN BD Topo). The coherence is very high as it is a single-pass acquisition.

can be easily improved by modifying the entries with no consequence on the merging approach.

The third step is an improvement step that is briefly detailed in Section V. The previously estimated DSM is projected onto the ground, and the layover and shadow are computed and compared with the classification. The edges of buildings are validated or corrected from this comparison. Some above-ground structures are also reclassified.

The algorithm is finally applied to a real data set (presented in Section II); the method and its accuracy are commented on in Section VI.

II. Data Take Description

The available data take is a single-pass interferometric SAR image acquired by RAMSES (ONERA SAR sensor) over Dunkerque (north of France). The X-band sensor was operated at submetric resolution. The baseline is about 0.7 m, which leads to an average ambiguity altitude of 180 m. The ambiguity altitude h_{amb} is computed from (2) with $\phi_{\mathrm{topo}} = 2\pi$: $h_{\mathrm{amb}} = (\lambda R \sin\theta / pB_{\perp})$.

The height accuracy δh depends on the phase standard deviation $\hat{\sigma}_{\phi}$ ($\delta h = (h_{\mathrm{amb}}/2\pi)\hat{\sigma}_{\phi}$). As a first approximation, $\hat{\sigma}_{\phi}$ is a function of the signal-to-noise ratio (SNR) and the number of looks L, i.e.,

$$\hat{\sigma}_{\phi} = \frac{\sqrt{1-\rho^2}}{\sqrt{2L}\rho} \qquad \rho = \frac{\mathrm{SNR}}{1+\mathrm{SNR}}. \tag{3}$$

Unfortunately, the theoretical SNR was not available; thus, $\hat{\sigma}_{\phi}$ has been estimated on a planar surface. It is about 0.1 rad, which leads to a height accuracy of about 2–3 m. This value is too high for a good DSM retrieval of small houses, but good results can be expected on large buildings.

Fig. 2 presents some extracts of this data set. The area is composed of large buildings (15 m high maximum) and residential parts with small houses. The global track also contains an industrial area with large buildings. In this paper, two districts (Bayard and the industrial area) have been selected to account for architectural diversity as much as possible.

An IGN BD Topo[1] is available on the area: this database gives building footprints (1-m resolution) and average height of building edges (1-m accuracy). Unfortunately, the lack of knowledge of SAR sensor parameters prevents us from registering the SAR data on the BD Topo precisely. Therefore, a manual comparison is performed between the estimated DSM and the BD Topo. This ground truth has been completed by an extensive visit of the place.

[1]Data take of the National Geographical Institute.

III. First-Level Processing

The initial input data are the amplitude of the SAR image, the interferogram, and the corresponding coherence. These three images are processed to get improved or higher-level information. In this section, six algorithms are proposed for this purpose. They are not claimed to be the most efficient to represent urban landscapes. Users may implement their own information extraction algorithms with no consequence on the fusion scheme. Therefore, we deliberately do not detail the algorithms at this stage; this paper is mainly dedicated to the merging part.

Most of the algorithms were developed especially for this study and were already published; the others are well-known methods, which are helpful to solve part of the problem. The readers can refer to the references for more details.

The operators that have been used in this paper can be divided in three groups.

1) Classification operator. A first classification, based on amplitude statistics, is computed [9]. The statistical model is a Fisher distribution. This model is dedicated to high-resolution SAR data over urban areas. The results are improved with the addition of coherence and interferogram [10]. The output is a classified image with seven classes (namely ground, dark vegetation, light vegetation, dark roof, medium roof, light roof/corner reflector, and shadow).
2) Filtering operator. The interferogram is filtered to remove global noise with an edge-preserving Markovian filtering [11]. It is a low-level operator that improves the information. The output is a filtered interferogram.
3) Structure extraction operators. Specific operators dedicated to the extraction of the main objects that structure the urban landscape (roads [12], corner reflectors [10], shadows, and isolated buildings extracted from shadow [13]) have been developed. The outputs are binary images (1 for the object sought after and 0 elsewhere).

Therefore, six new inputs (i.e., the filtered interferogram, the classification, the road map, the corner reflector map, the shadow map, and the building from shadow map) are now available from the three initial images. This new information is partly complementary and partly redundant. For instance, the corner reflectors are detected both with the dedicated operator and the classification. Generally speaking, the redundancy comes from very different approaches: the first one is local (classification), and the other one is structural (operators), accounting for the shape. This redundancy should lead to a better identification of these important structures.

IV. Fusion in a Markovian Framework

Starting from the six new inputs, our aim is to retrieve a height map and a classification with semantic classes. In some cases, only contextual information allows the retrieval of the

correct class of a pixel (for instance, roofs and trees may have close radiometries). Besides, this contextual information is not at the pixel level; small sets of pixels should be considered. In this case, two solutions can be developed: either the merging is conducted on large neighborhoods around each pixel, or it is conducted on small regions. The computational burden is larger in the first case. In addition, the neighborhoods do not preserve the shape for small objects. At this stage, the regions are determined easily from the new inputs without further computation; no additional computation cost is added. Therefore, it has been decided to consider regions rather than neighborhoods.

The regions and their neighborhoods are defined in Section IV-A. As a consequence, the DSM reconstruction issue becomes the recovery of height and urban object class for each region. The introduction of contextual knowledge between the regions uses a Markovian model that is defined in Sections IV-B and C. This assumption makes sense since a scene can be interpreted at a local scale by a photo-interpreter. The optimization algorithm, the parameters used, and their influence are addressed in Section IV-D.

A. Graph Definition

Some of the results computed in Section III are already based on contextual information: the classification operator (radiometric homogeneity) and the structure extraction operators (structural and radiometric/interferometric homogeneities). Therefore, these inputs are used for region definition. The boundaries of the classification and of the extracted objects (roads, etc.) are superimposed to define a partition of the scene. Each part of this "oversegmentation" is a region that will be considered as a node of the graph. The adjacency relationship is used to define the neighborhood of a region. A region adjacency graph (RAG) [14] is thus obtained, where each node is a region, and two nodes are linked if the corresponding regions are adjacent. As explained in the following paragraph, a characterizing vector is attached to each graph node. It contains the values of the different inputs and the surface of the node (i.e., the number of pixels per node).

B. Maximum A Posteriori (MAP) Formulation

In the following, bold characters are used for vectors. When possible, capitals are used for random variables and normal size characters for samples.

Two fields are defined on the RAG, namely: 1) the height field H and 2) the label field L. The height values are quantized to get discrete values from 0 to ambiguity altitude h_{amb} with a 1-m step. There is a small oversampling of the height regarding the expected precision. H_s, the random variable associated with node s, takes its value in $\mathbb{Z} \cap [0, h_{\mathrm{amb}}]$, and L_s takes its value in the finite set of urban objects: ground (G), grass (Gr), tree (T), building (B), corner reflector (CR), and shadow (S). These classes have been chosen to model all the main objects of towns as they appear in SAR images.

The six outputs in Section III define two fields $\bar{H}$ and $\mathbf{D}$ that are used as inputs of this merging step. $\bar{H}$ is the filtered interferogram and $\mathbf{D}$ is the observation field given by the classification and the structure extractions.

A value $\bar{h}_s$ of $\bar{H}$ for a region s is defined as the mean height of the filtered interferogram over this region. A value $\mathbf{d}_s = (d_s^i)_{1 \leq i \leq n}$ of $\mathbf{D}$ for a region s is defined as a vector built on the classification result and object extractions. This vector contains labels for the classification operator (here six classes are used) and binary values for the other operators (i.e., corner reflector, road, shadow, building estimated from shadows). They are still binary or "pure" classes because of the oversegmentation.

The aim is to subsequently find the configuration of the joint field (L, H) that maximizes the conditional probability $P(L, H|\mathbf{D}, \bar{H})$. It is the best solution using the MAP criterion. With the Bayes equation, we have

$$P(L, H|\mathbf{D}, \bar{H}) = \frac{P(\mathbf{D}, \bar{H}|L, H)P(L, H)}{P(\mathbf{D}, \bar{H})}$$

and

$$P(L, H) = P(L|H)P(H).$$

The joint probability is equal to

$$P(L, H|\mathbf{D}, \bar{H}) = \frac{P(\mathbf{D}, \bar{H}|L, H)P(L|H)P(H)}{P(\mathbf{D}, \bar{H})}. \quad (4)$$

Instead of supposing L and H independently, $P(L|H)$ is kept to constrain the class field by the height field. It usually allows one to take into account simple considerations on real architecture such as "roads are lower than adjacent buildings" or "herb and road are approximately at the same height." This link between H and L is the main originality and advantage of this approach.

Knowing the configurations $\mathbf{d}$ and $\bar{h}$, the denominator $(P(\mathbf{D}, \bar{H}))$ is constant and is not implied in the optimization of (L, H). Therefore, the final probability to be optimized is

$$P(L, H|\mathbf{D}, \bar{H}) = kP(\mathbf{D}, \bar{H}|L, H)P(L|H)P(H) \quad (5)$$

where k is a constant. The terms in (5) are defined in the following section.

C. Energy Terms

Assuming that both fields H and L conditionally dependent on H $(L|H)$ are Markovian, their probabilities are Gibbs fields. Adding the hypothesis of region-to-region independency conditionally dependent on L and H, the likelihood term $P(\mathbf{D}, \bar{H}|L, H)$ is also a Gibbs field. Indeed, $P(\mathbf{D}, \bar{H}|L, H) = \prod_s P(\mathbf{D}_s, \bar{H}_s|L, H)$, and assuming that the observation of regions does not depend on the other regions, $P(\mathbf{D}, \bar{H}|L, H) = \prod_s P(\mathbf{D}_s, \bar{H}_s|L_s, H_s)$. As a consequence, the energy is defined with a clique singleton. The posterior field is thus Markovian, and the MAP optimization of the joint field (L, H) is equivalent to the search for the configuration that minimizes its energy.

For each region s, the conditional local energy U is a function of the class l_s and the height h_s given the detector values $\mathbf{d}_s$, the observed height $\bar{h}_s$, and the field configuration of L and H of its neighborhood V_s. The energy is made up of two terms, namely: 1) the likelihood term U_{data} (coming from $P(\mathbf{D}, \bar{H}|L, H)$),

corresponding to the influence of the observations, and 2) the different contributions of the regularization term U_{reg} (coming from $P(L|H)P(H)$), corresponding to the prior knowledge that should be introduced on the scene. They are weighted by a regularization coefficient β and by the surface A_s of the region via a function α. The choice of the regularization terms (β and α) is empirical. The results do not change drastically with small (i.e., 10%) variations of β and α.

Taking into account the decomposition of the energy term into two energies (U_{reg} and U_{data}) and the weighting by the regularization term β and the surface function α, the following energy form is proposed:

$$U(l_s, h_s|\mathbf{d}_s, \bar{h}_s, l_t, h_{t\in V_s}) = (1-\beta)\left(\sum_{t\in V_s} A_t A_s\right)\alpha(A_s) \times U_{\text{data}}(\mathbf{d}_s, \bar{h}_s|l_s, h_s) + \beta \sum_{t\in V_s} A_t A_s U_{\text{reg}}(l_s, h_s, l_t, h_t) \quad (6)$$

where α is a linear function of A_s. If A_s is large, then the influence of the neighborhood is reduced ($\forall x, 1 \leq \alpha(x) \leq 2$). In addition, the different contributions of the regularization term are weighed by the surface product $A_t A_s$ to give more credit to the largest regions. The factor $(\sum_{t\in V_s} A_t A_s)$ is a normalization factor.

1) Likelihood Term: The likelihood term describes the probability $P(\mathbf{D}, \bar{H}|L, H)$. $\mathbf{D}$ and $\bar{H}$ are independent, thus $P(\mathbf{D}, \bar{H}|L, H) = P(\mathbf{D}|L, H) \times P(\bar{H}|L, H)$. Moreover, $\mathbf{D}$ is independent from H, and $\bar{H}$ is independent from L. Finally, $P(\mathbf{D}, \bar{H}|L, H) = P(\mathbf{D}|L) \times P(\bar{H}|H)$. Therefore, the likelihood term is considered equal to

$$U_{\text{data}}(\mathbf{d}_s, \bar{h}_s|l_s, h_s) = \sum_{i=1}^{n} U_D(d_s^i|l_s) + (h_s - \bar{h}_s)^2. \quad (7)$$

The likelihood term on the height is quadratic because of the Gaussian assumption over the interferometric phase probability [2]. There is no analytical expression of the density probability function of $P(d_s^i|l_s)$; it is thus determined empirically.

The values of $U_D(d_s^i|l_s)$ are determined by the user with respect to his *a priori* knowledge on the detector qualities. d_s^i values are part of finite sets (almost binary sets) because detectors deliver binary maps or classification. Therefore, the number of $U_D(d_s^i|l_s)$ values to be defined is not too high. Actually, d_s^1 is the classification operator result and has six possible values. The four others (d_s^2 the corner reflector map, d_s^3 the road map, d_s^4 the "building from shadow" map, and d_s^5 the shadow map) are binary map values. Therefore, the users have to define 96 values. Nevertheless, for binary maps, most of the values are equal because only one class is detected (the other ones are processed equally), which restricts the number of values to approximately 50. An example of the chosen values is given in Table I. To simplify the user choices, only eight values can be chosen: 0.0, 0.5, 0.8, 1.0, -3.0, -2.0, -10.0, and 3.0. Intermediate values do not have any impact on the results. The height map is robust toward changes of values, whereas the classification is more sensitive to small changes (from 0.8 to 0.5 for instance). Some confusion may arise between building and vegetation for such parameter changes.

TABLE I

$U_D(d_s^i|l_s)$ VALUES FOR EVERY CLASS AND EVERY DETECTOR. LINES CORRESPOND TO THE DIFFERENT VALUES THAT EACH ELEMENT d_s^i OF d_s HAVE, WHEREAS THE COLUMN CORRESPONDS TO THE DIFFERENT CLASSES CONSIDERED FOR l_s. EACH VALUE IN THE TABLE IS THUS $U(d_s^i|l_s)$ GIVEN THE VALUE OF d_s^i AND THE VALUE OF l_s. THE MINIMUM ENERGY VALUE IS 0.0 (MEANING "IT IS THE GOOD DETECTOR VALUE FOR THIS CLASS") AND THE MAXIMUM ENERGY VALUE IS 1.0 (MEANING "THIS DETECTOR VALUE IS NOT POSSIBLE FOR THIS CLASS"). THERE ARE THREE INTERMEDIATE VALUES, NAMELY: 1) 0.3; 2) 0.5; AND 3) 0.8. YET WHEN SOME DETECTORS BRING OBVIOUSLY STRONG INFORMATION, WE UNDERLINE THEIR ENERGY BY USING ± 2, ± 3, OR -10 REGARDING THE CONFIDENCE LEVEL. IN THIS WAY, CORNER REFLECTOR AND SHADOW DETECTORS ARE ASSOCIATED WITH THE LOW ENERGY BECAUSE THESE DETECTORS CONTRIBUTE TRUSTWORTHY INFORMATION THAT CANNOT BE CONTESTED. THE MERGING IS ROBUST REGARDING SMALL VARIATION OF ENERGY VALUES. CR = CORNER REFLECTORS, R = ROADS, BS = BUILDINGS FROM SHADOWS, S = SHADOWS, B = BUILDING, S = SHADOW. THE CLASSIFICATION VALUES d_s^1 MEAN 0 = GROUND, 1 = VEGETATION, 2 = DARK ROOF, 3 = MEAN ROOF, 4 = LIGHT ROOF, AND 5 = SHADOW. THE CLASSES ARE GROUND (G), GRASS (GR), TREE (T), BUILDING (B), CORNER REFLECTOR (CR), AND SHADOW (S)

d_s^i \ l_s		G	Gr	T	B	CR	S
Classification	$d_s^1=0$	0.0	1.0	1.0	1.0	1.0	1.0
	$d_s^1=1$	1.0	0.0	0.8	1.0	1.0	1.0
	$d_s^1=2$	1.0	0.5	0.0	0.0	1.0	1.0
	$d_s^1=3$	1.0	1.0	0.5	0.0	1.0	1.0
	$d_s^1=4$	1.0	1.0	1.0	0.0	0.0	1.0
	$d_s^1=5$	1.0	1.0	1.0	1.0	1.0	-3.0
CR	$d_s^2=0$	1.0	1.0	1.0	1.0	3.0	1.0
	$d_s^2=1$	1.0	1.0	1.0	1.0	-2.0	1.0
R	$d_s^3=0$	1.0	1.0	1.0	1.0	1.0	1.0
	$d_s^3=1$	-10.0	1.0	1.0	1.0	1.0	1.0
BS	$d_s^4=0$	0.0	0.0	0.3	0.5	0.0	0.0
	$d_s^4=1$	1.0	1.0	0.3	0.0	0.3	1.0
S	$d_s^5=0$	1.0	1.0	1.0	1.0	1.0	3.0
	$d_s^5=1$	1.0	1.0	1.0	1.0	1.0	-2.0

Moreover, these values are defined once over the entire data set and are not modified regarding the particularities of the different parts of the global scene.

2) Regularization Term: The contextual term, relating $P(L|H)P(H)$, introduces two constraints and is written as

$$U_{\text{reg}}(l_s, h_s, l_t, h_t) = \gamma_{(h_s,h_t)}(l_s, l_t) + \psi(h_s - h_t). \quad (8)$$

The first term γ comes from $P(L|H)$ and imposes constraints on two adjacent classes l_s and l_t depending on their heights. For instance, two adjacent regions with two different heights cannot belong to the same road class. A set of such simple rules is built up and introduced in the energy term.

The second term ψ comes from $P(H)$ and introduces contextual knowledge on the reconstructed height field. Since there are many discontinuities in urban areas, the regularization should both preserve edges and smooth planar regions (ground, flat roof).

For the class conditionally dependent on heights, the adjacency of two regions is encouraged or discouraged regarding relative height difference. Three cases have been distinguished, namely: 1) $h_s \approx h_t$; 2) $h_s < h_t$; and 3) $h_s > h_t$, and an adjacency matrix is built for each case. To preserve symmetry, the matrix of the last case is equal to the transposed matrix of the second case.

At $h_s \approx h_t$, we have

$$\gamma_{(h_s,h_t)}(l_s, l_t) = 0 \text{ if } (l_s, l_t) \in \{B, CR, S\}^2 \quad (9)$$

$$\gamma_{(h_s,h_t)}(l_s, l_t) = \delta(l_s, l_t) \text{ else} \quad (10)$$

where δ is the Kronecker symbol.

TABLE II
$c(l_s, l_k)$ VALUES, I.E., $\gamma_{(h_s,h_k)}(l_s, l_k)$ VALUES WHEN $h_s < h_k$. THE SYMMETRIC MATRIX GIVES THE VALUES OF $\gamma_{(h_s,h_k)}(l_s, l_k)$ WHEN $h_s > h_k$. FOUR VALUES ARE USED FROM 0.0 TO 2.0. 0.0 MEANS THAT IT IS HIGHLY PROBABLE TO HAVE CLASS l_s CLOSE TO CLASS l_k, WHEREAS 2.0 MEANS THE EXACT CONTRARY (IT IS ALMOST IMPOSSIBLE). THE CLASSES ARE GROUND (G), GRASS (GR), TREE (T), BUILDING (B), CORNER REFLECTOR (CR), AND SHADOW (S)

l_s \ l_k	G	Gr	T	B	CR	S
G	1.0	2.0	0.5	0.5	2.0	1.0
Gr	2.0	1.0	0.5	0.5	2.0	1.0
T	2.0	2.0	0.0	1.0	2.0	1.0
B	1.0	1.0	1.0	0.0	0.0	0.0
CR	2.0	2.0	2.0	0.0	0.0	1.0
S	1.0	1.0	1.0	0.0	1.0	0.0

In this case, the two adjacent regions have similar height and should belong to the same object. Yet, if the region is a shadow or a corner reflector, the height may be noisy and could be close, in average, to that of the building.

At $h_s < h_t$, we have

$$\gamma_{(h_s,h_t)}(l_s, l_t) = c(l_s, l_t). \tag{11}$$

At $h_s > h_t$, we have

$$\gamma_{(h_s,h_t)}(l_s, l_t) = c(l_t, l_s). \tag{12}$$

These last two cases relate the real relationship between classes regarding their height. The user has to define the values $c(l_s, l_t)$ regarding real town structure. But there is a unique set of values for an entire data set. An example of the chosen values is given in Table II.

For the height, regularization is calculated with an edge-preserving function [11]

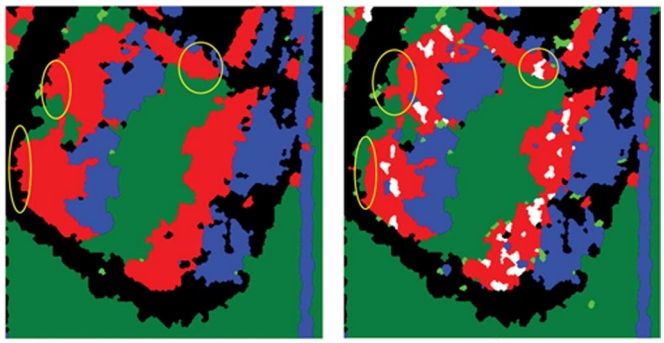
$$\psi(h_s, h_t) = \frac{(h_s - h_t)^2}{1 + (h_s - h_t)^2}. \tag{13}$$

This function is a good compromise to keep sharp edges while smoothing planar surfaces.

D. Optimization Algorithm

Due to computational constraints, the optimization is processed with an iterative conditional mode (ICM) algorithm [15]. The classification initialization is computed from the detector inputs. The maximum-likelihood value is assigned to the initial class value, i.e., for each region, the initial class l_s is the one that minimizes $\sum_{i=1}^{n} U_D(d_s^i | l_s)$. The initialization of the height map is the filtered interferogram. This initialization is close to the expected results, which allows an efficient optimization through the ICM method.

The algorithm is run with specific values: the regularization coefficient β is given a value of 0.4, and the α function is equal to $\alpha(A) = (A - \min(A_s))/(\max(A_s) - \min(A_s)) + 1$. $\min(A_s)$ and $\max(A_s)$ are, respectively, the minimum and maximum region surfaces of the RAG. The energy terms defined by the user are presented in Tables I and II. These values are used for the entire data take; they are not adapted to each extract. For a given data set, the user has thus to define these values only once.



(a) (b)

Fig. 3. Illustration of the improvement step. The ellipses simply underline three areas with major improvement. Some edges are corrected (for instance, the two regions circled on the left), and some missing corner reflectors are added (the region circled on the right). (a) Initial classification. (b) Improved one. In the two cases on the left, some vegetation has been classified as building due to its amplitude property. Yet, its height is small, and the computation of the layover part proves that these regions cannot belong to a building. In the case on the right, no corner reflector has been detected, but, as it is a building, it should appear somewhere. The correction step enables us to compute the position of the corner reflector and to add it in the classification.

V. IMPROVEMENT STEP

The final step will correct some errors in the classification and DSM by checking the coherency between them. In this part, two RAGs are considered, namely: 1) the one defined for the merging step (based on regions) and 2) a new one constructed from the final classification l. The regions of the same class, in the first graph, are gathered to obtain the complete object, leading to an object adjacency graph.

The corrections are performed for each object. When an object is flagged as misclassified, it is split in regions again (according to the previous graph) to correct only the misclassified parts of the objects.

The correction steps include the following.

- Rough projection of the estimated DSM on ground geometry.
- Computation of the "layover and shadow map" from the DSM in ground geometry (ray tracing technique).
- Comparison of the estimated classification with the previous map l, detection of problems (for instance, layover parts that lay on ground class or layover parts that do not start next to a building).
- Correction of errors. For each flagged object, the partition of regions is reconsidered, and the region not compliant with the layover and shadow maps is corrected. For layover, several cases are possible: if layovers appear on ground regions, the regions are corrected as trees or buildings depending on their size; for buildings that do not start with a layover section, the regions in front of the layover are changed into grass. Height is not modified at this stage.

Thanks to this step, some building edges are corrected, and missing corner reflectors are added. The effects of the improvement step on the classification are illustrated in Fig. 3. The comparison of layover start and building edges allows the edges to be relocated. In some cases, the building edges are

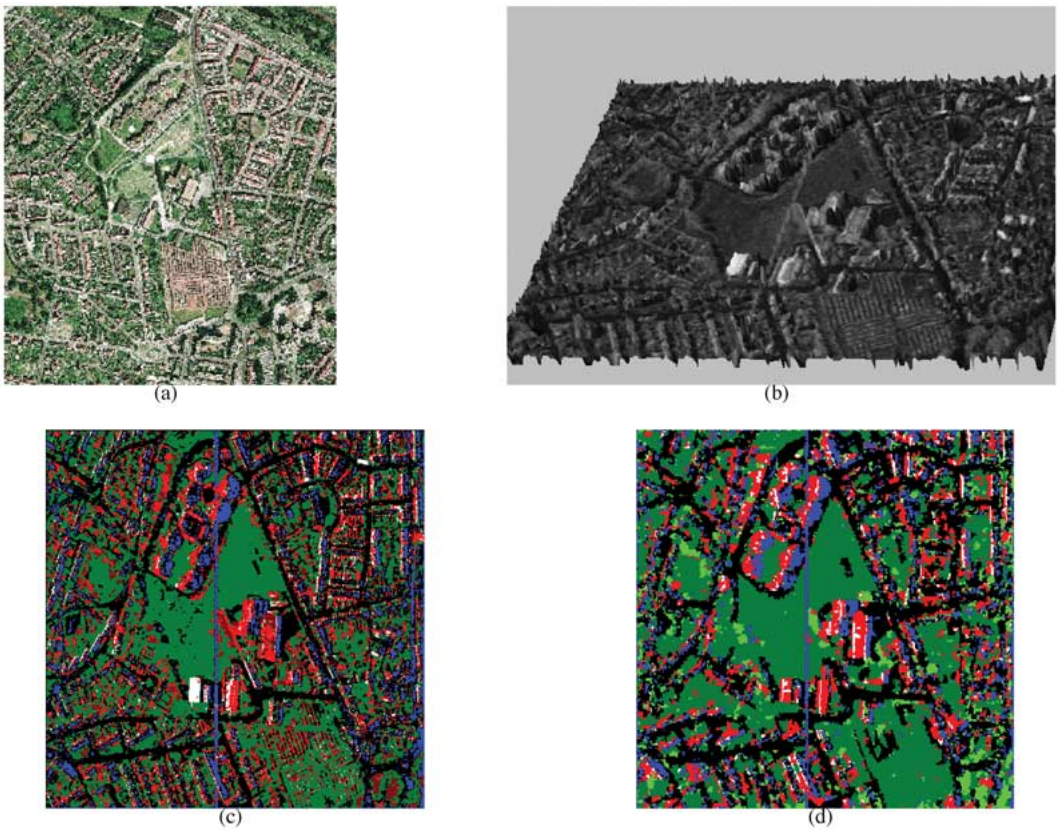


Fig. 4. Results of the Bayard district. (a) Optical image (IGN). (b) Three-dimensional view of the DSM with SAR amplitude image as texture. (c) Classification used as input. (d) Final classification. (Black = streets, dark green = grass, light green = trees, red = buildings, white = corner reflector, blue = shadow.)

mispositioned due to small objects close to the edges. They are discarded through layover comparison.

In the very last step, the heights of vegetation regions are reevaluated: it does not make sense to have a mean value for a region of trees. Thus, the heights of the filtered interferogram are kept in each pixel (instead of a value per region). Actually, tree regions do not have a single height, and the preservation of the height variations over these regions enables us to stay closer to reality.

VI. Application on Real Data

The fusion scheme presented in this paper has been tested on a real set of high-resolution interferometric data (presented in Section II). Two districts in Dunkerque have been selected for their architectural diversity, namely: 1) the Bayard district and 2) an industrial area. The Bayard district includes a large panel of buildings (isolated buildings, residential areas, and straight and curved roads), whereas the industrial area features large metallic buildings with strong backscattering.

The energy terms have been defined only once for the entire X-band Dunkerque data set, and the values in Tables I and II are used for both extracts.

A. Results on the Two Test Sites

The results are presented in Fig. 4 (Bayard district) and Fig. 5 (Industrial area), where the global understanding of the scene appears to be very good with regards to the altimetric criterion and to the class criterion. The first point is the good retrieval of the road map in the Bayard district scene (the industrial case does not present a real road map). The backbone of the urban areas is thus well estimated. Then, the global shape of major buildings is realistic, even for complex buildings. For instance, the large building in the upper left part of the Bayard district is very complex with several parts, irregular shapes, and different roof elevations. Yet the estimated DSM and the classification are very close to reality: the estimated elevations of the inner courtyard and the roofs are right even if all the small details of the roof are lost.

The height map is well regularized for flat areas, while some roof details are present thanks to the region approach. For instance, roof arches of the building on the middle right part of the industrial area are kept (Fig. 6). This proves that the roofs are not all modeled by flat surfaces. Such levels of detail are only available on large buildings. In the case of small houses with a sloping roof, the resolution is too poor to actually estimate correctly the two slopes. The method will provide them with an estimate if higher resolution data are available.

Nevertheless, due to poor altimetric precision (2–3 m, see Section II), small gaps of less than 2 m appear on the flat surface, such as roads or grass. They are due to altimetric noise and should not be considered as information. A height sampling rate equal to the noise height value will provide smoother results.

The classification result is not corrupted by height noise, and the final result is clearly better than the classification obtained at the first step. The fusion scheme solves ambiguities between

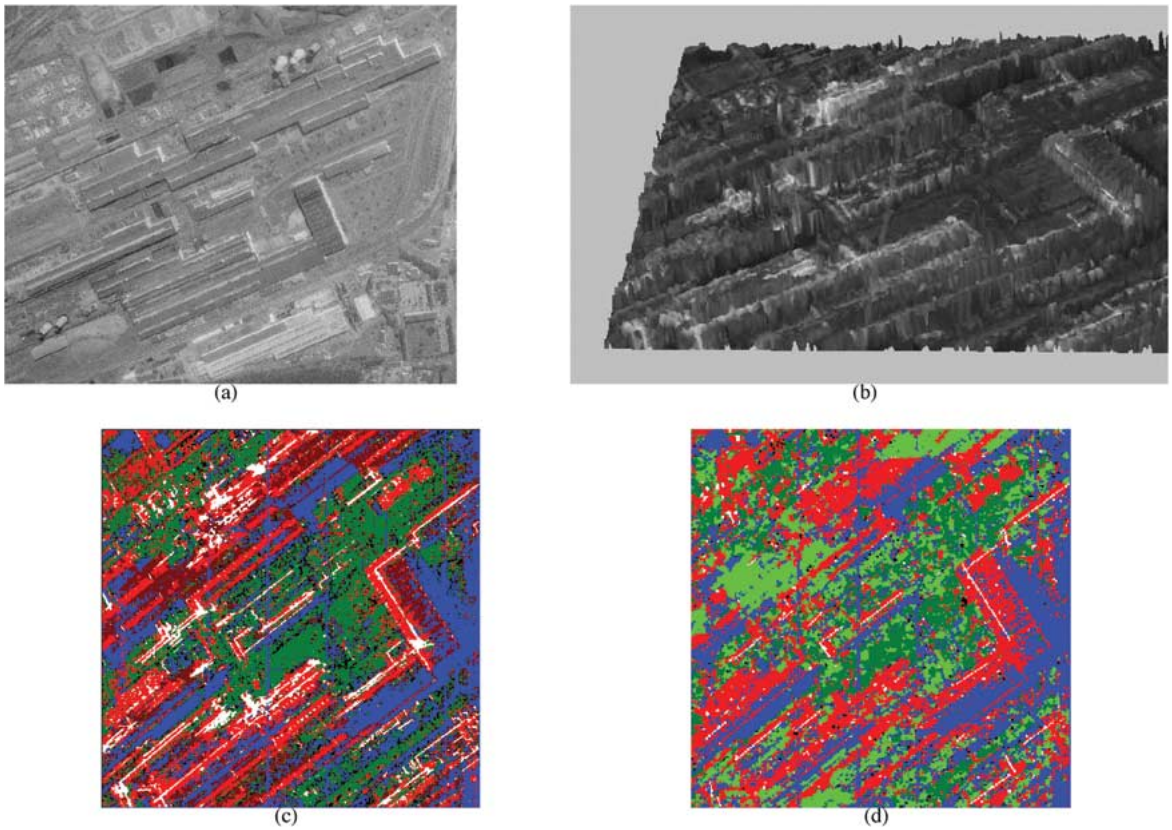


Fig. 5. Results of the industrial area. (a) Optical image (IGN). (b) Three-dimensional view of the DSM with SAR amplitude image as texture. (c) Classification used as input. (d) Final classification. (Black = streets, dark green = grass, light green = trees, red = buildings, white = corner reflector, blue = shadow.)

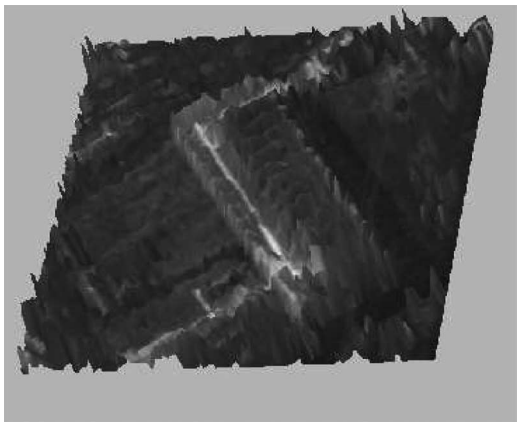


Fig. 6. Roof detail of the reconstructed industrial scene. The different arches of the roof are very well reconstructed from the interferogram.

trees and buildings (some holes in building roofs are filled in). Yet some confusion remains between trees and buildings as their statistical properties may be very close. The detectors that have been used do not take into account geometrical shapes, which is here the only means to separate buildings and trees when the proposed approach fails.

In addition, the classification is not accurate on very small structures (such as residential areas) because the spatial resolution is too low. For instance, the small houses on the lower left part of the Bayard district are not well retrieved because they are too small. The classification acknowledges the presence of man-made structures, but the edges are very approximate.

B. Comparison With Ground Truth

A manual comparison between ground truth and estimated DSM has been conducted on 19 buildings of the Bayard area. They have been picked out to describe a large variety of buildings (small and large ones, regular and irregular shapes). The mean height of the estimated building is compared with the mean height of the BD Topo (ground truth). For each building, the estimated height is plotted versus the expected height (Fig. 7). A perfect estimation will lead to results close to the $(y = x)$ line. A small deviation can be observed. Actually, the root mean square error is around 2.5 m, which is the best result that could be expected in view of the altimetric precision (2–3 m).

C. Critical Analysis

First, altimetric and spatial image resolutions have a very strong impact on the quality of the result. They cannot be ignored for result analysis. From these results, the spatial resolution has to be better than 50 cm and the altimetric precision better than 1 m to preserve all the structures for a very accurate reconstruction of dense urban areas (containing partly small houses). When these conditions are not met, one should expect poor quality results on the smallest objects, which can be observed with our data set. This conclusion is not linked with the reconstruction method.

Second, a typical confusion is observed for every scene: buildings and trees are not always well differentiated. They both present similar statistical properties and can only be

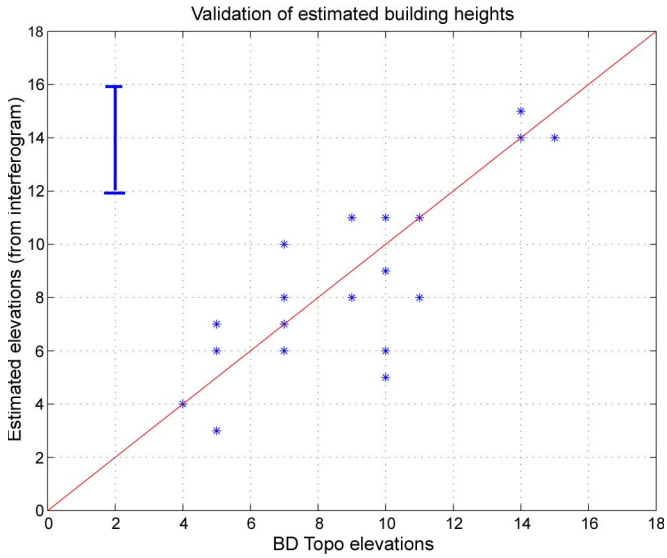


Fig. 7. Comparison of building mean height estimated from interferogram versus building mean height of IGN BD Topo over 19 buildings of Bayard district. Each star corresponds to a building. The error bar in the upper left part of the plot represents the uncertainty associated with each interferometric measurement. This uncertainty issues from the height accuracy of the interferogram (see Section II).

differentiated with their geometry. In fact, building shape is expected to be very regular (linear or circular edges, right angles, etc.) compared with vegetation areas (at least in towns). A solution may be the inclusion of geometrical constraints to discriminate buildings from vegetation. Stochastic geometry is a possible investigation field to add a geometrical constraint after the merging step.

This problem appears mostly for the industrial areas where there are no trees. In this case, the user may add an extra information in the algorithm (for instance suppression of the tree class) to reach a better result. This has been successfully tested. This example proves that an expert will get better results than a novice, or a fully automatic approach. Actually, the complexity of the algorithm and the data requires expertise. The user has to fix some parameters for the merging step level (energy, weighting values). Nevertheless, once the parameters have been assigned for a given data set, the entire data set can be processed with these values. Yet locally some extra information may be required, e.g., a better selection of the class.

Nevertheless, the method remains very flexible: users can change detection algorithms or energy terms to improve the final results without altering the processing chain architecture. For instance, the detection of shadows is not optimum yet, and better detection will certainly improve the final result.

VII. Conclusion

The purpose of this paper is to complete the processing chain for retrieving DSM over urban areas from high-resolution SAR interferogram. Emphasis is put on the merging step, where classification and DSM are retrieved jointly. The mutual relations between class and height are used to improve both products.

The results are very promising: the estimated heights are close to the real ones when building sizes are large enough with regards to image and altimetric resolutions. In addition, the global shape of the buildings, roads, and trees (namely, the structure of the town) is well retrieved. Of course, results are less convincing on residential areas as resolutions are too coarse in this context. Good results can reasonably be expected in such situations when a finer resolution is available. Nevertheless, higher resolutions may infer new properties of the SAR signal, which cannot be ignored.

The method presented here can be easily improved by modifying the entries of the merging step. The fusion scheme is completely independent of the meaning and the number of these entries.

Another important point that should be addressed is the confusion between vegetation and building. In some cases, the only discriminating feature is the shape. Stochastic geometry may thus be a good approach to solve the ambiguity. It could be initialized by the DSM and the classification to reduce computational costs.

As a conclusion, SAR interferometry proves to be a relevant method to compute DSM over urban areas. Some developments are still necessary to obtain an operational processing chain. A deep need for such a chain may arise because new SAR missions are about to be launched (TerraSAR-X, CosmoSkymed, SAR Lupe) and will provide a large amount of interferometric images. In this context, one can expect to get series of interferometric couples of the same area. The combination of several interferograms over the same scene will surely improve the final results. In particular, shadows and layovers may be better accounted for in a multi-image context. This study is the first step for a more general use of interferograms, and the results should be considered as encouraging for future work in this field.

Acknowledgment

The authors would like to thank CNES (DCT/SI/AR) and EADS DCS (LTIS), especially to J.-C. Souyris (CNES) and V. Leroy (LTIS), for their support, and the Office National d'Etudes et de Recherches Aérospatiales and the Délégation Générale pour l'Armement for providing the data.

References

[1] D. Massonnet and T. Rabaute, "Radar interferometry: Limits and potentials," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 2, pp. 455–464, Mar. 1993.

[2] P. Rosen, S. Hensley, I. Joughin, F. Li, S. Madsen, E. Rodrìguez, and R. Goldstein, "Synthetic aperture radar interferometry," *Proc. IEEE*, vol. 88, no. 3, pp. 333–382, Mar. 2000.

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

[4] 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.

[5] M. Quartulli and M. Dactu, "Information extraction from high resolution SAR data for urban scene understanding," in *Proc. 2nd GRSS/ISPRS Joint Workshop—'Data Fusion and Remote Sens. Over Urban Areas'*, May 2003, pp. 115–119.

[6] D. Petit, "Reconstruction du 3-D par interférométrie radar haute résolution," Ph.D. dissertation, IRIT, Toulouse, France, Jan. 2004.

[7] U. Soergel, U. Thoennessen, and U. Stilla, "Iterative building reconstruction from multi-aspect InSAR data," in *Proc. ISPRS Working Group III/3 Workshop*, Oct. 2003, vol. XXXIV, pp. 186–192.

[8] F. Tupin, I. Bloch, and H. Maître, "A first step toward automatic interpretation of SAR images using evidential fusion of several structure detectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1327–1343, May 1999.
[9] C. Tison, J. Nicolas, F. Tupin, and H. Maître, "A new statistical model of urban areas in high-resolution SAR images for Markovian segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2046–2057, Oct. 2004.
[10] C. Tison, F. Tupin, and H. Maître, "Extraction of urban elevation models from high resolution interferometric SAR images," in *Proc. EUSAR*, Ulm, Germany, May 2004, pp. 411–414.
[11] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.
[12] G. Lisini, C. Tison, F. Tupin, and P. Gamba, "Feature fusion to improve road network extraction in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 2, pp. 217–221, Apr. 2006.
[13] C. Tison, F. Tupin, and H. Maître, "Retrieval of building shapes from shadows in high resolution SAR interferometric images," in *Proc. IGARSS*, Sep. 2004, vol. 3, pp. 1788–1791.
[14] J. W. Modestino and J. Zhang, "A Markov random field model-based approach to image interpretation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 6, pp. 606–615, Jun. 1992.
[15] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Stat. Soc.*, vol. 48, no. 3, pp. 259–302, 1986.

Céline Tison received the Engineering degree and the Ph.D. degree from the Ecole Nationale Superieure des Telecommunications (Telecom Paris), Paris, France, in 2001 and 2004, respectively.

She is currently with the Radar and Altimetry Department, Centre National d'Etudes Spatiales, Toulouse, France. Her main research interests are SAR images, in particular, high resolution, urban areas, interferometry, and elevation extraction.

Florence Tupin received the Engineering degree and the Ph.D. degree from the Ecole Nationale Superieure des Telecommunications (Telecom Paris), Paris, France, in 1994 and 1997, respectively.

She is currently an Associate Professor with the Signal and Image Processing Department, Telecom Paris. Her main interests are image analysis and interpretation, Markov random-field techniques, and SAR remote sensing.

Henri Maître received the Engineering degree from the Ecole Centrale de Lyon, Ecully, France, in 1971 and the Docteur es Sciences degree in physics from the University of Paris VI, Paris, France, in 1982.

Since 1973, he has taught digital picture processing at the Ecole Nationale Superieure des Telecommunications (Telecom Paris), Paris, where he is currently a Professor. He is Head of the Signal and Image Processing Department and the Director of the LTCI Laboratory, Centre National de la Recherche Scientifique, Paris. His research includes work on image analysis, image understanding, and computer vision, and applications in the domains of satellite and aerial image processing and processing of documents issued from the fine arts.