TELECOM PARIS
école nationale supérieure des télécommunications

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

Un schéma de fusion pour l'optimisation conjointe de la classe et de la hauteur en interférométrie radar haute résolution

---

Céline Tison
Florence Tupin
Henri Maître

2006C002

juin 2006

Département Traitement du Signal et de l'Image
Groupe Traitement et Interprétation des Images

Dépôt légal : 2006 – 3ème trimestre
Imprimé à l'Ecole Nationale Supérieure des Télécommunications – Paris
ISSN 0751-1337 ENST C (Paris) (France 1983-9999)

Ecole Nationale Supérieure des Télécommunications
Groupe des Ecoles des Télécommunications - membre de ParisTech
46, rue Barrault - 75634 Paris Cedex 13 - Tél. + 33 (0)1 45 81 77 77 - www.enst.fr
Département TSI

Un schéma de fusion pour l'optimisation conjointe de la classe et de la hauteur en interférométrie radar haute résolution

Céline Tison**, Florence Tupin*, Henri Maître*
* GET-Télécom Paris - CNRS UMR 5141
46 rue Barrault, 75013 Paris, France
** CNES, DCT/SI/AR
18 avenue E. Belin, 31 40à Toulouse, France

Un des enjeux majeurs de la télédétection est la reconstruction tri-dimensionnelle de la Terre. Pour des surfaces naturelles et des résolutions moyennes, une grande partie du globe a déjà été imagée grâce aux missions SRTM ou SPOT HRS. Un nouveau défi est celui de l'obtention de MNS (Modèle Numérique de Surfaces) sur les zones urbaines. L'amélioration récente de la résolution des images radar à ouverture synthétique (SAR), et l'intérêt de l'interférométrie qui a fait ses preuves sur des scènes naturelles à basse résolution, font de l'imagerie radar un outil de premier plan pour la reconstruction 3D en milieu urbain. Pourtant, la complexité des zones urbaines et les mécanismes de rétrodiffusion propres à l'imagerie radar rendent nécessaires les traitements des données interférométriques avant de pouvoir accéder à un MNS. Dans cet article, nous proposons une méthode originale pour résoudre ce problème.

La chaîne de traitements que nous présentons se décompose en trois grandes étapes : extraction d'information, fusion, et correction. Notre contribution principale concerne l'étape de fusion, dans laquelle nous calculons simultanément une classification et le MNS associé. Cette étape conjointe nous permet d'introduire des informations contextuelles et des règles sur l'architecture des scènes réelles, tout en préservant un grande souplesse sur la forme des bâtiments cherchés.

Tout d'abord, les données interférométriques initiales (phase, amplitude et cohérence) sont converties en informations de plus haut niveau par différentes approches (filtrage, reconnaissance d'objets, classification) pour avoir une première interprétation de la scène. Dans un deuxième temps, ces nouvelles données sont fusionnées dans un cadre Markovien pour obtenir de façon conjointe une classification et une carte des hauteurs. Finalement, le MNS et la classification sont corrigés et améliorés par estimation des zones d'ombres et de repliements.

Ce papier détaille essentiellement l'étape de fusion Markovienne ; les deux autres étapes sont brièvement expliquées et font référence à des articles précédemment publiés.

Les résultats, obtenus sur des images réelles, sont comparés à une vérité terrain et montrent une bonne précision compte tenu de la résolution altimétrique et planimétrique des données originales.

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

Céline Tison*, Florence Tupin†, Henri Maître†

† GET-Télécom Paris - CNRS UMR 5141

46 rue Barrault - 75013 Paris - France

* CNES, DCT/SI/AR

18 avenue Edouard Belin

31 401 Toulouse cedex 4 - France

Phone number : + 33 5 61 28 24 04 – Fax number : + 33 5 61 28 34 22

Email : celine.tison@cnes.fr

Abstract

A major issue for remote sensing is the retrieval 3D surface from Earth, which is already available for natural surfaces at medium resolution, thanks to missions such as SRTM or SPOT HRS. A new challenging issue is now the derivation of DSM (Digital Surface Model) over urban areas. Since the recent improvement of radar image resolution, SAR (Synthetic Aperture Radar) interferometry which has proved its efficiency for natural scenes at low resolution may provide an accurate tool for urban 3D monitoring. However, the complexity of urban areas and high resolution SAR images prevents from computing an accurate DSM straightforwardly. In this article, we propose an original high level processing chain to solve this problem and we present some results on real data.

The processing chain includes three main steps: information extraction, fusion and correction. Our main contribution addresses the merging step, where we aim at retrieving both a classification and a DSM while imposing less constraints as possible on the building shapes. The joint derivation of height and class enables to introduce more contextual information and rules describing real scenes and, thus, to be more flexible towards scene geometry. 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) to get a first understanding of the scene. Secondly, these new images are merged into a Markovian framework to retrieve jointly an improved classification and a height map. Thirdly, DSM and classification are improved by computing layovers and shadows from the estimated DSM. Comparison between shadows/layovers and classification allows some corrections.
This paper mainly addresses on the second step; the two others are briefly explained and referred to already published articles.

The results (obtained on real images) are compared to ground truth and indicate a very good accuracy in spite of limited image resolutions. The major limit of the DSM computation remains the adequacy of the initial spatial and altimetric resolutions.

Index Terms

SAR interferometry, urban areas, Markovian fusion, height map, classification

I. INTRODUCTION

The extraction of 3D town models is a major issue for many applications, such as environment or urban planning. Thanks to recent improvement of SAR (Synthetic Aperture Radar) image resolution, SAR interferometry can now address this issue. Future SAR missions (SAR Lupe, CosmoSkymed, TerraSAR-X) will deliver high resolution interferometric data with global coverage. As a consequence, the evaluation of interferometry potential over urban areas is a subject of most concern. This paper presents an original and flexible method to extract DSM (Digital Surface Model) from high resolution interferogram over urban areas. Due to the complexity of such areas, a dedicated scheme is required.

We deliberately restrict ourselves to the use of one single interferometric data take per scene in order to fully assess the interferometry potential. Our challenge was also to develop a method with no restriction on building shapes, so that every city type is expected to be compliant with this technique.

A. Interferometry and urban areas context

The interferometry is based on the phase difference of two SAR images acquired over the same scene with a slightly different incidence angle. Under some 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). In order to reduce noise, we introduce an averaging over a $L \times$ Ł size window:

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

ϕ has two contributions: an orbital one ϕ_{orb}, linked to natural line-of-sight vector variation across the scene and a topographical one ϕ_{topo}. By Taylor expanding to first order, the height h in every pixel is proportional to ϕ_{topo} and depends on wavelength λ, sensor-target distance R, perpendicular baseline $B_\perp$ and incidence angle θ:

$$h = \frac{\lambda}{4\pi} \frac{R \sin\theta}{B_\perp} \phi_{topo} \tag{2}$$

ϕ_{orb} is only geometry dependent and can easily be removed from ϕ [2]. Therefore, in the following, interferometric phase should be understood as topographic phase only (orbital phase was removed previously).

Although Eq. 2 looks simple, the direct inversion does not lead to an accurate Digital Surface Model (DSM). The first reason is the knowledge of the phase modulo 2π that requires a phase unwrapping. This topic is not addressed in this paper because ambiguity altitude is high enough compared to building heights.

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

B. State of the art

High resolution SAR images remain quite new and, as a consequence, available for a small community only. Therefore, literature on DSM retrieved from SAR interferometry is only at its beginning. So far, four kinds of methods have been proposed:

- shape from shadow [3]:

  Building footprints are estimated from the shadows detected in the amplitude image, whereas the interferogram provides only 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 strong and is not very realistic in the context of spaceborne data takes.
- roof filtering [4]:

  The interferogram gives a noisy height map which is filtered out by looking for horizontal planes (roof

buildings) initialized by 3D segments.

These first two methods are only efficient for large and isolated buildings.

- stochastic geometry [5]:

  The position and the shape of each building is optimized by a function linking the amplitude, the coherence and the interferogram. Because of computational time constraint, the possible building shape is restricted to a unique model. As a consequence, a strong a priori is made on urban architecture.

- global scene reconstruction based on a primarily classification [6], [7]:

  This approach is the most flexible and operational (at least [7] reference). It links 3D reconstruction and classification and does not require any shape assumption. Nevertheless the significant results obtained by Soergel et al. [7] relies on the merging of several interferograms over the same scene and on a rectangular shape model for buildings.

The first original aspect of our work is that the input data over a scene are deliberately limited to an interferometric couple and that no constraint on building shapes are considered. Actually, in operational context, the user may have only one interferogram per scene. In addition, the town architecture diversity is generally important: it cannot be restricted to one building model. This framework made us select the fourth approach. However, instead of dealing with the classification and with DSM estimation separately, we propose a joint computation of height and class. In fact, class and height have strong interactions that should be taken into account to improve the global scene recovery.

C. Proposed method

The global processing is divided into three main steps (Fig. 1). Since the original SAR data are difficult to interpret, new inputs are preliminary derived from pattern recognition methods, denoising, classification, etc. to get higher level information. This step is briefly described in Section III and mainly refers to previous works. In addition, the algorithms proposed for this step and the associated results should be considered only as open options. Users are free to develop their own tools to derive first step information. This will not impact the global architecture of the processing.

In a second step, all these new images are merged into a Markovian framework to provide jointly a classification and a height map. The merging method is inspired of [8]. First an over segmentation of the scene is applied to define regions, on which classification and height recovery will be applied. This region partitioning allows the reduction of computation time and the accounting for 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 observation field (Section IV). The global architecture of this second step is completely independent on the number and content of the inputs. Therefore the result 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 ground and the layovers and shadows are computed and compared to the classification. From this comparison, the edges of buildings are validated or corrected. Some above ground structures are reclassified.

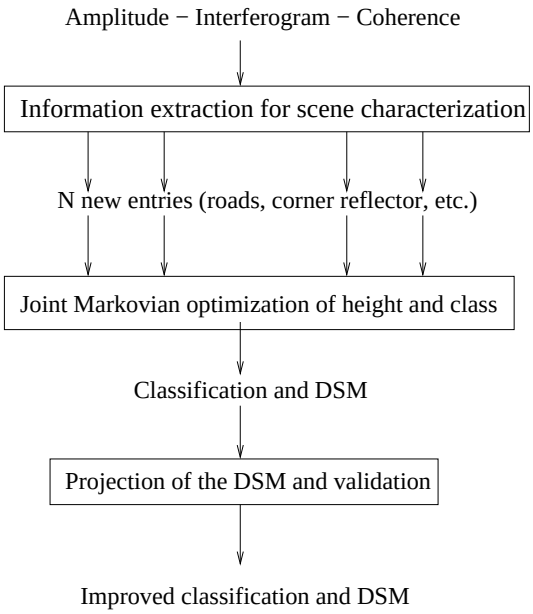


Fig. 1. Global scheme of the proposed method for DSM estimation over urban areas. The height estimation is processed jointly with a classification, as these two pieces of information are deeply linked.

The algorithm is finally applied to a real dataset (presented in Section II) and the method and the result quality are commented (Section VI).

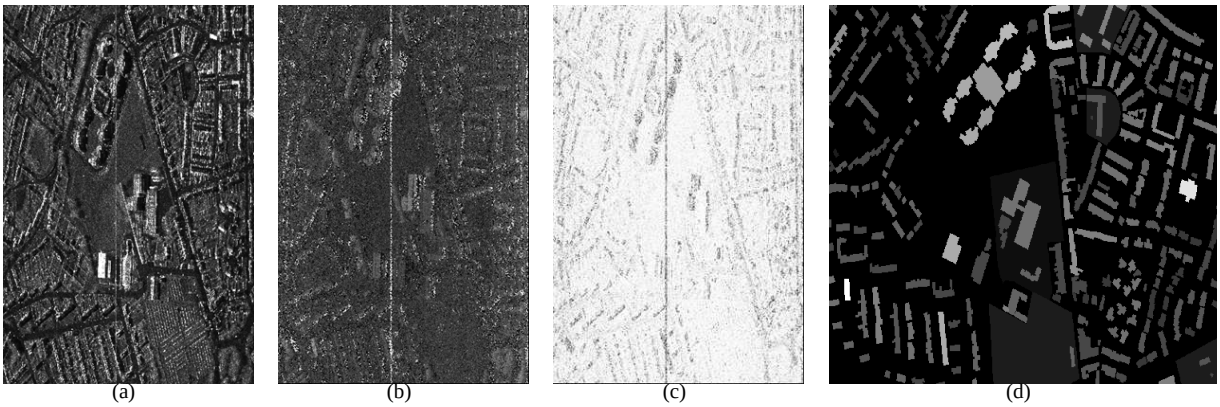
II. Data take description

The available data take is single pass interferometric SAR images acquired by RAMSES (ONERA SAR sensor) over Dunkerque (North of France). The X-band sensor was operated at sub-metric resolution. The baseline is about 0.7 m, which leads to an average ambiguity altitude of 180 meters. This value does not permit very accurate height estimation. The altimetric accuracy is about 2-3 meters regarding the ambiguity altitude and the noise level. At this stage, we already know that DSM computation over small houses will fail but we can expect good results for large buildings.

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

An IGN BD Topo©[1] is available on the area: this database gives building footprints (one meter resolution) and average height of building edges (one meter accuracy). Unfortunately, the lack of knowledge on SAR sensor parameters prevents from registering the SAR data on the BD Topo© precisely. Therefore, we performed a manual

[1]Data take of the national geographical institute.



(a) (b) (c) (d)

Fig. 2. Presentation of the available data take (Bayard district): (a) multi-look amplitude, (b) interferogram, (c) coherence and (d) ground truth (IGN BD Topo©). The coherence is very high as it is a single pass acquisition.

comparison between the estimated DSM and the BD Topo©. This ground truth has been completed by an extensive visit of the place.

III. First level processings

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, we propose six algorithms to come up with it. We do not claim them to be the most efficient to represent the urban landscapes at a first step. The 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 as the 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 helpful to solve part of the problem. The readers can refer to the references for more details. The used operators can be divided in three groups:

- classification operator:

  A first classification is computed based on amplitude statistics [9]. The statistical model is a Fisher distribution and is dedicated to high resolution SAR data over urban areas. The results are improved by adding coherence and interferometric information [10]. The output is a classified image with seven classes (ground, dark vegetation, light vegetation, dark roof, medium roof, light roof/corner reflector and shadow);
- filtering operator:

  the interferogram is filtered to remove global noise with an edge preserving Markovian filtering [11]; it is a low-level operator which gives an improved information. The output is a filtered interferogram;
- structure extraction operators:

  specific operators dedicated to the extraction of the main objects structuring 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 dedicated object, 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. These new information are partly complementary and partly redundant. For instance, the corner reflectors are detected both with the dedicated operator and the classification. Generally speaking, the redundancy issues 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 for retrieving 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 set up: either the merging is conducted on large neighbourhoods around each pixel, or it is conducted on small regions. The computational burden is larger in the first case. In addition, the neighbourhoods do not preserve 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, we decided to consider regions rather than neighbourhoods.

The region definition and their neighbourhoods are described in Part IV-A. As a consequence DSM reconstruction issue becomes the recovery issue of height and urban object class for each region. The introduction of contextual knowledge between the regions is made by using a Markovian model which is defined in Parts IV-B and IV-C. This assumption makes sense since the interpretation of a scene can be done at a local scale by a photo-interpreter. The optimization algorithm, the used parameters and their influence are addressed in Part IV-D.

A. Graph definition

Some of the results computed in Section III are already based on contextual information: the classification operator (radiometric homogeneity), the structure extraction operators (structural and radiometric or interferometric homogeneity). Therefore these inputs of the fusion part 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 “over-segmentation” is a region which will be considered as a node of the graph. The adjacency relationship is used to define the neighbourhood 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. The region surface is added to the characteristics of its graph node.

B. Maximum a posteriori formulation

(In the following, bold characters are used for vectors. When possible, capital characters are used for random variables and normal size characters for realizations.)

Two fields are defined on the RAG: the height field H and the label field L. The height values are quantified in order to get discrete values from 0 to ambiguity altitude h_{amb} with a one meter step. There is a small oversampling of the height regarding the expected precision. The realization h_s of H for node s takes its value into $\mathbb{Z} \cap [0, h_{amb}]$ and the realization l_s of L takes its value into the finite set of urban objects: {Ground (G), Grass (Gr), Tree (T), Building (B), Corner Reflector (CR), Shadow (S)}. These classes have been chosen to model all the main objects of towns as they appear in SAR images.

The six outputs of Section III define two fields $\bar{H}$ and $\mathbf{D}$, that are used as inputs of this merging step. $\bar{H}$ corresponds to the filtered interferogram which will play a key-role in the following and $\mathbf{D}$ corresponds to the observation field constituted of the classification and the structure extractions.

A realization $\bar{h}_s$ of $\bar{H}$ for a region s is defined as the mean height of the filtered interferogram over this region. A realization $\mathbf{d}_s = (d_s^i)_{1 \leq i \leq n}$ 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 over-segmentation.

The aim is subsequently to find the realization of the joint field (L, H) which maximizes the conditional probability $P(L, H|\mathbf{D}, \bar{H})$. It is the best solution using a Maximum A Posteriori (MAP) criterion. With the Bayes equation:

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

The joint probability is equal to:

$$P(L, H|\mathbf{D}, \bar{H}) = \frac{P(\mathbf{D}, \bar{H}|L, H)P(L|H)P(H)}{P(\mathbf{D}, \bar{H})} \tag{3}$$

Instead of supposing L and H independent, $P(L|H)$ is kept to constrain the class field by the height field. It usually allows 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.

The field L is assumed to be equiprobable (thus the prior $P(L)$ is constant) and regularization is only processed on H and $L|H$. Therefore the final probability to be optimized is:

$$P(L, H|\mathbf{D}, \bar{H}) = kP(\mathbf{D}, \bar{H}|L, H)P(L|H)P(H) \tag{4}$$

with k a constant. Terms of Eq. 4 are defined in the following section.

C. Energy terms

Assuming that both fields L conditionally to H and H are Markovian, their probabilities are Gibbs fields. Adding the hypothesis of region to region independency conditionally to 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, $P(\mathbf{D}, \bar{H}|L, H) = \prod_s P(\mathbf{D}_s, \bar{H}_s|L_s, H_s)$ which will lead to an energy with clique

singleton. The posterior field is thus Markovian and the MAP optimization of the joint field (L, H) is equivalent for the search of the configuration that minimizes its energy.

For each region s, the conditional local energy U is function of the class l_s and the height h_s knowing the detector values $\mathbf{d}_s$, the observed height $\bar{h}_s$ and the field configuration of L and H of its neighbourhood V_s. The energy is made of two terms: the likelihood term U_{data} (coming from $P(\mathbf{D}, \bar{H}|L, H)$) corresponding to the influence of the observations, and the different contributions of the regularization term U_{reg}(coming from $P(L|H)P(H)$) corresponding to the prior knowledge we wish to introduce on the scene. They are weighted by a regularization coefficient β and by the surface A_s of the considered region via a function α. The choice of the regularization terms (β and α) are empirical. The results do not change drastically with small (i.e. 10%) variations of β and α.

We propose the following energy form:

$$U(l_s, h_s|\mathbf{d}_s, \bar{h}_s, l_t, h_{t t\in V_s}) = (1-\beta)(\sum_{t\in V_s} A_t A_s)\alpha(A_s)U_{data}(d_s, \bar{h}_s|l_s, h_s) + \beta \sum_{t\in V_s} A_t A_s U_{reg}(l_s, h_s, l_t, h_t) \quad (5)$$

α is a linear function of A_s. If A_s is large then the influence of the neighbourhood is reduced ($\forall x, 1 \leq \alpha(x) \leq 2$). In addition, the different contributions of the regularization term is weighted by the surface product $A_t A_s$ in order 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 is taken equal to:

$$U_{data}(d_s, \bar{h}_s|l_s, h_s) = \sum_{i=1}^{n} U_D(d_s^i|l_s) + (h_s - \bar{h}_s)^2 \quad (6)$$

The values of $U_D(d_s^i|l_s)$ are determined by the user regarding 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 stands for 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 maps. Therefore the users have to define ninety-six values. Nevertheless, for binary maps, most of the values are equal, because only one class is detected (the other one are treated equally), which restricts the number of values to approximately fifty. 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 and -3.0, -2.0, -10.0 and 3.0. Intermediate values do not have any impacts on the results. The height map is robust towards changes of values whereas the classification is more sensible to small changes (from 0.8 to 0.5 for instance). For instance, confusion arises between building and vegetation for such parameter modifications.
Moreover these values are defined once for all (over the entire data set), but are not modified regarding the particularities of the different parts of the global scene.

The likelihood term on the height is quadratic because of the Gaussian assumption over the interferometric phase probability [2].

2) Regularization term: The contextual term introduces two constraints and is written in Eq. 7. The first one 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 are built up and taken back in the energy term.

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

$$U_{reg}(l_s, h_s, l_t, h_t) = \gamma_{(h_s,h_t)}(l_s, l_t) + \psi(h_s - h_t) \tag{7}$$

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

$h_s \approx h_t$:

$$\gamma_{(h_s,h_t)}(l_s, l_t) = 0 \text{ if } (l_s, l_t) \in \{B, CR, S\}^2 \tag{8}$$

$$\gamma_{(h_s,h_t)}(l_s, l_t) = \delta(l_s, l_t) \text{ else} \tag{9}$$

δ is the Kronecker symbol.

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

$h_s < h_t$:

$$\gamma_{(h_s,h_t)}(l_s, l_t) = c(l_s, l_t) \tag{10}$$

$h_s > h_t$:

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

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 heights, the regularization is made with an edge preserving function [11]:

$$\psi(h_s, h_t) = \frac{(h_s - h_t)^2}{1 + (h_s - h_t)^2} \tag{12}$$

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

D. Optimization algorithm

Due to computational constraints, the optimization is processed with an ICM algorithm [15]. The classification initializing is computed from the detector inputs as the maximum likelihood, i.e. for each region, the initial class l_s is the one which 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 ICM method.

The algorithm is run with specific values detailed here. The regularization coefficient β is taken equal to 0.4; the α function is equal to $\alpha(A) = \frac{A-\min(A_s)}{\max(A_s)-\min(A_s)} + 1$. $\min(A_s)$ and $\max(A_s)$ are, respectively, the minimum and the maximum region surfaces of the RAG. The energy terms defined by the user are presented in Tab. 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.

TABLE I

$U_D(d_s^i|l_s)$ VALUES FOR EVERY CLASS AND EVERY DETECTORS. 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: 0.3, 0.5 AND 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 TO LOW ENERGY BECAUSE THESE DETECTORS BRING TRUSTFUL INFORMATION THAT SUFFERS NO CONTEST. 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, 5 = SHADOW.

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

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).

l_i \ l_j	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

V. IMPROVEMENT STEP

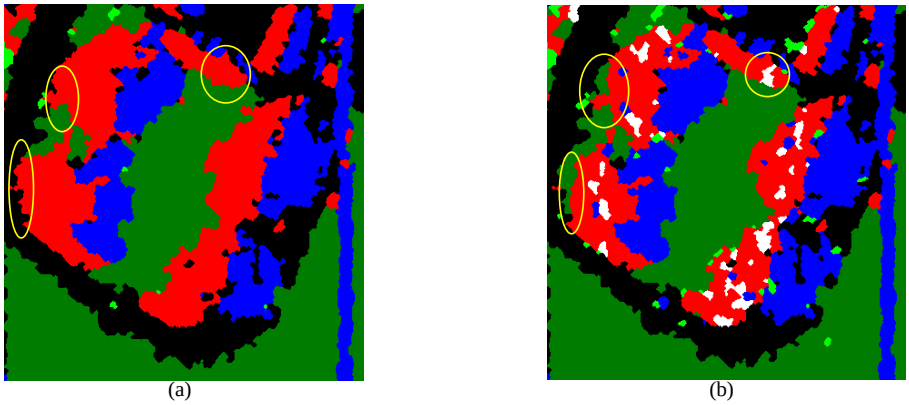
The final step will correct some errors in classification and DSM by checking the coherency between the two results. In this part, two region adjacency graphs are considered: the one defined for the merging step (based on regions) and a new one constructed from the final classification l. The regions of same class 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 miss-classified, it is split in regions again (according to the previous graph) in order to correct only the misclassified parts of the objects.

The main steps include:

- 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 region partition is reconsidered and the region not compliant with layover and shadow map is corrected. For layovers, several cases are possible: if layovers appear on ground regions, the regions are corrected as trees or buildings depending on their size; for building that do not start with a layover section, the regions in front of the layover are changed into grass. The 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 over the classification are illustrated on Figure 3. The comparison of layover start and building edges allows relocating the edges. In some cases, the building edges are mispositionned due to small objects close to the edges. They are discarded through layover comparison.



(a) (b)

Fig. 3. Illustration of the improvement step. Some edges are corrected and some missing corner reflectors are added. The initial classification is (a) and the improved one is (b). The ellipses simply underline three areas with major improvement.

In a very last step, the heights of vegetation region are re-evaluated: it does not make sense to have a mean value over a region for trees. Then 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 to stay closer to reality.

VI. Application on real data

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

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

A. Results on the two test sites

The results are presented in Fig. 4 and Fig. 5, where the global understanding of the scene appears to be very good, with regards to the altimetric criterion and to the class criterion. The height map is well regularized for flat areas, while some roof details are presented thanks to the region approach. For instance, roof arches are kept (Fig. 6).

Due to a poor altimetric precision (2-3 meters, see Section II), small gaps of less than 2 meters appear on flat surface, such as roads or grass. They are due to the altimetric noise and should not be considered as information. A height sampling rate equal to the noise height value will enable to get 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 enables to solve some ambiguities between trees and buildings (some holes in building roofs are filled in). Yet some confusion remains between trees and buildings as their statistical

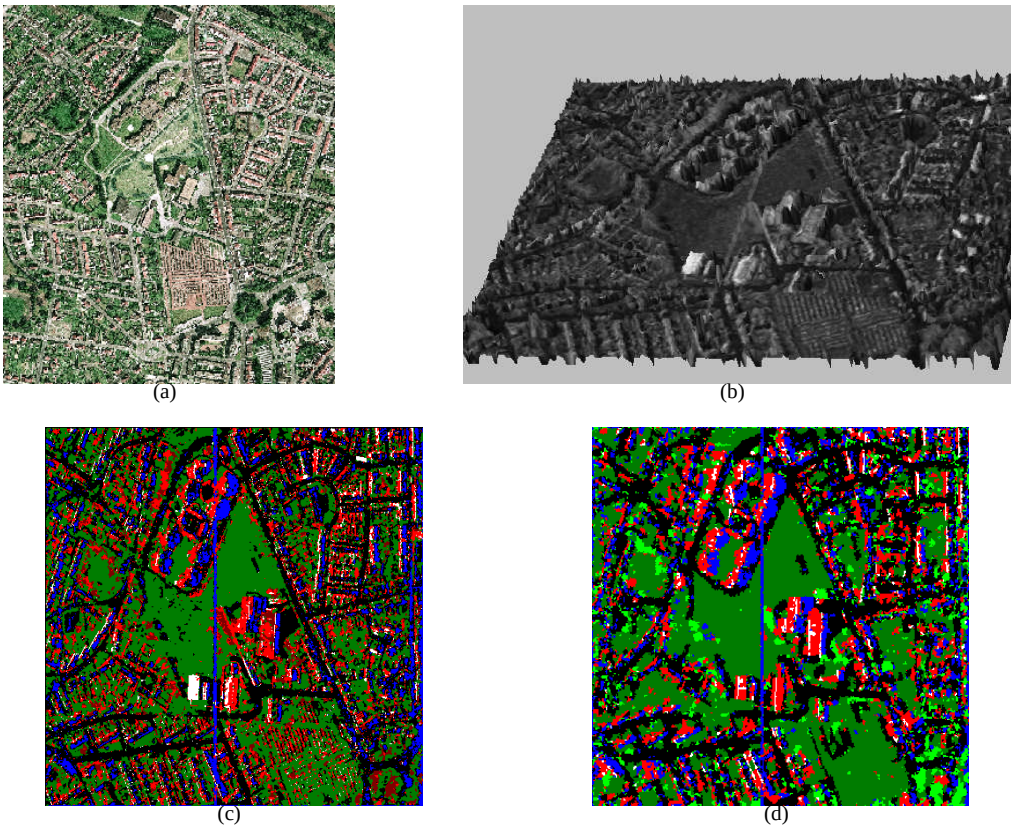


Fig. 4. Results of the Bayard district: (a) optical image (IGN), (b) 3D view of the DSM with SAR amplitude image as texture, (c) classification used as input, (d) final classification. (black=streets, dark green=grass, light green=trees, red=buildings, white=corner reflector, blue=shadow)

properties may be very close. In addition, the classification is not accurate on very small structures (such as residential areas) because spatial resolution is too low.

B. Comparison with ground truth

A manual comparison between ground truth and estimated DSM has been conducted on several buildings: the mean height of the estimated map is compared to the mean height of the BD Topo©. The rms error is around 2.5 meters (Fig. 7), which is the best result that can be expected given the altimetric precision (2-3 m).

C. Critical analysis

Firstly, altimetric and spatial image resolutions have a very strong impact on result quality. They cannot be ignored for result analysis. From these results, we can assume that, for a very accurate reconstruction of dense urban areas (containing partly small houses), the spatial resolution has to be better than 50 cm and the altimetric precision better than 1 meter to preserve all the structures. When these conditions are not met, one should expect bad quality results on the smaller objects, which can be observed in the dataset presented here. This conclusion is independent from the reconstruction method.

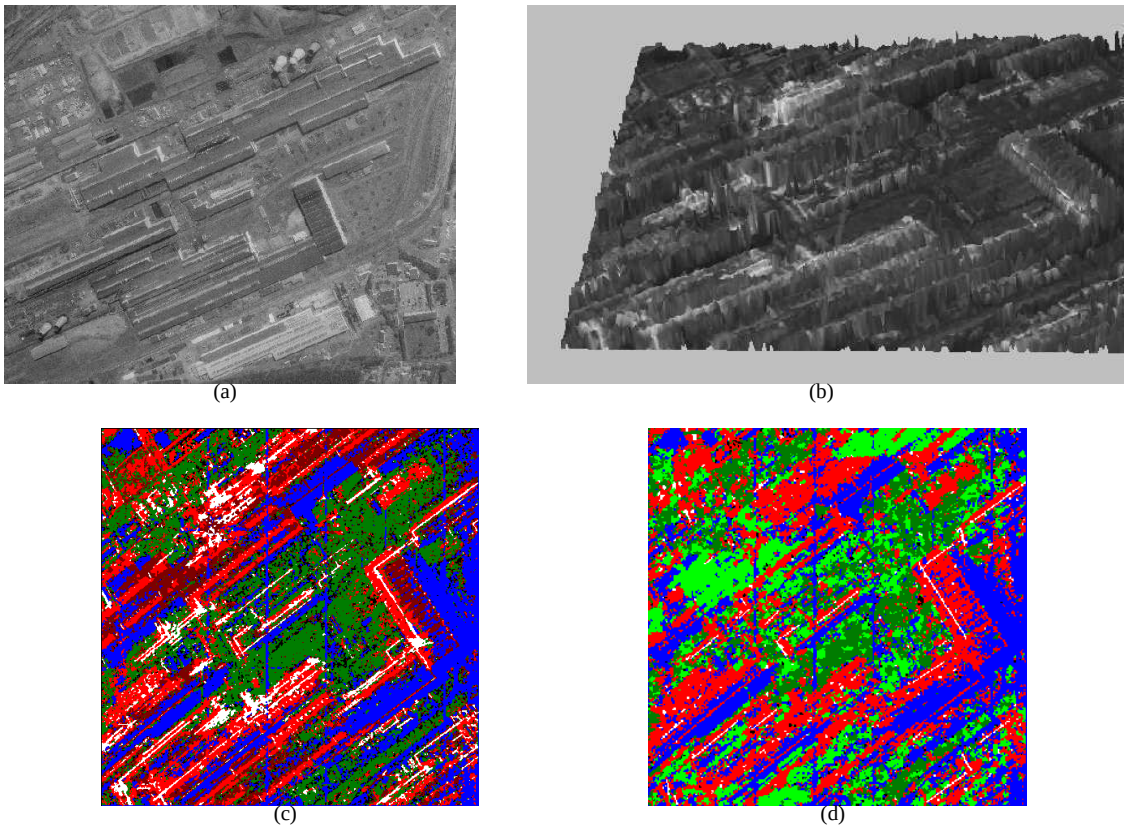


Fig. 5. Results of the industrial area: (a) optical image (IGN), (b) 3D view of the DSM with SAR amplitude image as texture, (c) classification used as input, (d) final classification. (black=streets, dark green=grass, light green=trees, red=buildings, white=corner reflector, blue=shadow)

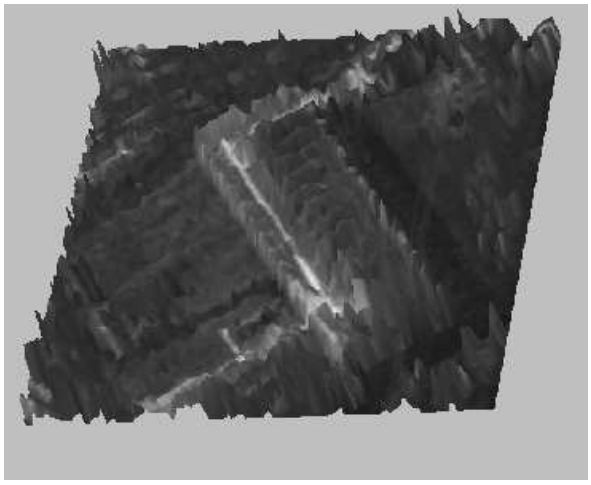


Fig. 6. Roof detail of the reconstructed industrial scene. The different arches of the roof are very well reconstructed from the interferogram.

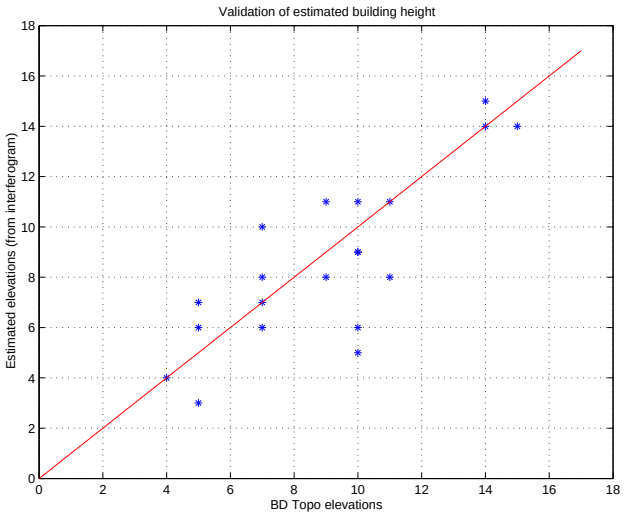


Fig. 7. Comparison of building mean height estimated from interferogram with building mean height of IGN BD Topo© over some buildings of Bayard district.

Secondly, a typical confusion is observed for every scene: buildings and trees are not always well differentiated. They both present similar statistical properties; their main difference is the 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 geometrical constraint after the merging step.

In the results, this problem appears mostly in the industrial areas where there is no tree. In this case, the user may add an extra-information in the algorithm (i.e. suppression of the tree class) to reach better result. This has been successfully tested. This example proves that an expert will get better result than a novice, or than a fully automatic approach. Actually the complexity of the algorithm requires expertise as it is not fully automatic. The user has to fix some parameters at the merging step level (energy, weighting values). Nevertheless once the parameters have been assigned for a given data set, the entire dataset can be processed with these values. Yet locally some added information may be required, e.g. a better selection of the class that should be fine in the scene.

Nevertheless the method remains very flexible: users can change detection algorithms or energy terms to improve the final results without altering the processing architecture.

VII. Conclusion

The purpose of this article was to complete processing chain for retrieving DSM over urban areas from high resolution SAR interferogram. Emphasize is put on the merging step, where a classification and a 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, the roads and

the 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. We can reasonably expect good results over in such situations when finer resolution will be available. Nevertheless higher resolutions may infer new properties of 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 on the meaning and the number of these entries. For instance, the detection of shadows is not optimum yet and better detection will certainly improve the final result.

Another important point to address is confusion issue between vegetation and building. In some cases, the only discriminate 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, in order to reduce computational costs, even by allowing several building models.

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 but it is worth it as many high resolution interferometric images should be available in the near future (TerraSAR-X, CosmoSkymed, SARLuppe). In this context, one can expect to get series of interferometric couples of the same area that will surely improve the final results in comparison with a single interferometric acquisition. In particular, shadows and layovers may be better accounted in multi-images context. This study is a first step for a more general use of interferograms and the results should be considered as encouraging for future works on this field.

ACKNOWLEDGMENT

The authors are indebted to CNES (DCT/SI/AR) and EADS DCS (LTIS), especially to Jean-Claude Souyris (CNES) and Vincent Leroy (LTIS) for their support. They also thank ONERA[2] and DGA[3] for providing the data.

REFERENCES

[1] D. Massonnet and T. Rabaute, “Radar interferometry: Limits and potentials,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, pp. 445–464, 1993.

[2] P. Rosen, S. Hensley, I. Joughin, F. Li, S. Madsen, E. Rodrìguez, and R. Goldstein, “Synthetic aperture radar interferometry,” in *Proceedings of the IEEE*, vol. 88, no. 3, 2000, pp. 333–382.

[3] R. Bolter, “Reconstruction of man-made objects from high resolution SAR images,” in *IEEE Aerospace Conference*, vol. 3, Mar. 2000, pp. 287–292.

[4] P. Gamba, B. Houshmand, and M. Saccani, “Detection and extraction of buildings from interferometric SAR data,” *IEEE Trans. on Geosc. and Rem. 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 *2nd GRSS/ISPRS Joint Workshop on ”Data Fusion and Remote Sensing over Urban Areas”*, May 2003, pp. 115–119.

[6] D. Petit, “Reconstruction du ”3D” par interférométrie radar haute résolution.” Ph.D. dissertation, IRIT, Jan. 2004.

[7] U. Soergel, U. Thoennessen, and U. Stilla, “Iterative building reconstruction from multi-aspect InSAR data,” in *ISPRS working group III/3 Workshop*, vol. XXXIV, Oct. 2003.

[2]ONERA=Office National d’Etudes et de Recherches Aérospatiales

[3]DGA=Délégation Générale pour l’Armement

[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 Transactions on Geoscience and Remote Sensing*, 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. on Geosc. and Rem. Sens.*, Oct. 2004.

[10] C. Tison, F. Tupin, and H. Maître, "Extraction of urban elevation models from high resolution interferometric SAR images," in *EUSAR 2004*, Ulm, Germany, May 2004, pp. 411–414.

[11] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. on PAMI*, vol. 14, no. 3, pp. 367–383, 1992.

[12] G. Lisini, C. Tison, D. Cherifi, F. Tupin, and P. Gamba, "Improving road network extrcation in high resolution SAR images by data fusion," in *CEOS*, Ulm, Germany, May 2004.

[13] C. Tison, F. Tupin, and H. Maître, "Retrieval of building shapes from shadows in high resolution SAR interferometric images," in *IGARSS'04*, vol. III, Sept. 2004, pp. 1788–1791.

[14] J. W. Modestino and J. Zhang, "A Markov Random Field model-based approach to image interpretation," *IEEE Trans. on PAMI*, vol. 14, no. 6, pp. 606–615, 1992.

[15] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc.*, pp. 259–302, 1986.