To link to this article: DOI: 10.1109/TGRS.2011.2113351
URL : http://dx.doi.org/10.1109/TGRS.2011.2113351

High-Resolution Optical and SAR Image Fusion for Building Database Updating

Vincent Poulain, Jordi Inglada, *Member, IEEE*, Marc Spigai,
Jean-Yves Tourneret, *Senior Member, IEEE*, and Philippe Marthon

***Abstract*—This paper addresses the issue of cartographic database (DB) creation or updating using high-resolution synthetic aperture radar and optical images. In cartographic applications, objects of interest are mainly buildings and roads. This paper proposes a processing chain to create or update building DBs. The approach is composed of two steps. First, if a DB is available, the presence of each DB object is checked in the images. Then, we verify if objects coming from an image segmentation should be included in the DB. To do those two steps, relevant features are extracted from images in the neighborhood of the considered object. The object removal/inclusion in the DB is based on a score obtained by the fusion of features in the framework of Dempster–Shafer evidence theory.**

***Index Terms*—Change detection, data fusion, feature extraction, image analysis, image databases (DBs).**

I. Introduction

With the recent (or in the very next future) availability of high-resolution (HR) optical and radar satellite sensors, such as in the ORFEO program,[1] the need of multisensor image processing chains that are able to assist a human expert in scene interpretation is increasing.

In this work, we focus on the problem of cartography creation/update, and more precisely, on built-up areas. We propose a generic image processing and interpretation chain for cartography creation/update. This chain is generic because it can process multisensor data [optical and synthetic aperture radar (SAR) images are considered in this work] at various resolutions (between 70 cm and 2.5 m for results presented in this paper) and can take into account ancillary data (typically a digital map). HR optical images are often used in cartographic applications, thanks to their easy interpretation. However, optical sensors are time and weather dependent. On the contrary, SAR sensors have an equal effectiveness at any time of the day and night. They can quickly provide information in emergency situations or in cloudy area for instance. However, the interpretation of SAR images is more complex. Consequently, a multisensor application can exploit the complementarity of sensors to provide a maximum of information on a scene.

Several scenarios are possible according to available images and data. The basic (and optimistic) case is the one where available input data are the following: 1) a multispectral HR image (for instance, Quickbird or Pleiades); 2) an HR SAR image (for instance, Cosmo-SkyMed or TerraSAR-X); and 3) a vector database (DB). The aim is then to update the vector DB. However, other (less optimistic) scenarios of input data are foreseen to be processed by the chain: for instance, the use of a single optical or SAR image, or two images of the same kind (Cosmo-Skymed and TerraSAR-X for instance), or an optical image at a lower resolution (SPOT-5 for instance) or panchromatic, with or without DB as prior information.

Often, existing methods are specific to one sensor in single mode. Indeed, in the field of building extraction with a single optical image, many methods have been proposed. In [1] and [2], hypotheses of buildings are created by grouping primitives extracted from airborne images. Buildings are extracted in [3] from a panchromatic QuickBird image using clustering and edge detection. Methods based on segmentation of HR images followed by a segment classification to detect buildings are presented in [4] with multispectral images, in [5] with panchromatic images, and in [6] and [7] with aerial RGB images. Reference [8] presents a method based on active contours to check a digital map of buildings using a panchromatic Quickbird image. Methods based on a shape prior, using morphological operators, are presented in [9]–[11]. A building detection method based on the integration of shape priors in a level-set image segmentation is proposed in [12]. Approaches described in [13] and [14] use graph theory to extract buildings from optical images.

With SAR sensors, the analysis of a single image to extract buildings is a more challenging task. Some promising methods are based on marked point processes [15], [16]. However, robust results are very hard to achieve. Building detection and height estimation methods are proposed in [17]–[19] using interferometric SAR data. A method for extracting building outlines using a SAR intensity image is explained in [20], and using features extracted from SAR and optical images in [21]. In dense urban environments, single SAR images are more efficiently used to extract the road network. For instance, in [22], a road network extraction is proposed based on a Hough transform and a road tracking algorithm.

Contrary to these approaches, this paper proposes a generic chain. It is able to integrate multisensor images and exogenous data. The goal is to exploit all the available information on a scene. The proposed chain can also evolve with the easy integration of new features.

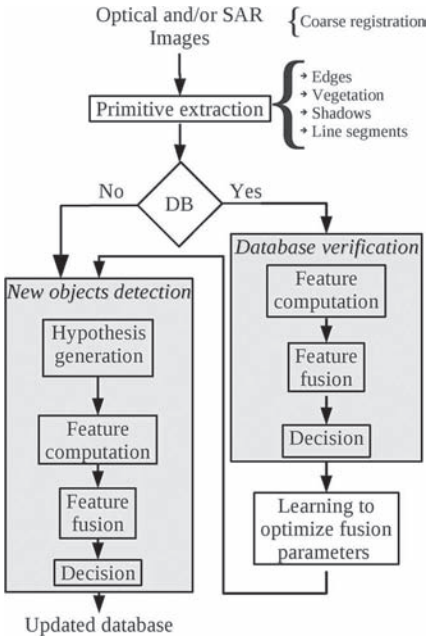


Fig. 1. Processing chain for building DB creation or updating.

This paper is organized as follows. Section II gives an overview of this work by describing our generic processing chain. Section III describes features used to characterize the objects of interest (buildings). A fusion method allowing one to combine all advantages of these features is studied in Section IV. Finally, results are presented in Section V.

II. Processing Chain

Our processing chain is presented in Fig. 1. Inputs of the chain are HR images, i.e., an optical and/or a SAR image. Resolution of images should be in the range of 0.6–2.5 m for the optical images, and around 1 m for the SAR images. A cartographic DB can be available. The registration of images and DB is not part of this work. Thus, images and DB are assumed to be registered. As we work at object level with buffer regions, a coarse registration (that can be performed automatically with an accuracy of several pixels) is sufficient to define appropriate buffer regions. Consequently, a fine registration is not needed. The goal of this chain is to update (if available) or create a vector DB representing buildings. If a DB is available, our approach consists of two steps: First, we consider each DB object and we check if the object is present in the SAR and optical images. To do so, some relevant features based on primitives are computed in the neighborhood of each object and fused to decide if the object should be kept in the DB. The second step consists of detecting buildings that are missing in the DB and including them in the DB. The proposed method is similar to the first step. However, instead of considering each DB object, we consider each region coming from a multiscale segmentation [23] of the optical image. In this work, an optical image is required to extract new buildings. If no optical image is available, the detection of new buildings should be performed with approaches specific to building detection in SAR images such as [15] and [16]. However, these methods are not used in this work.

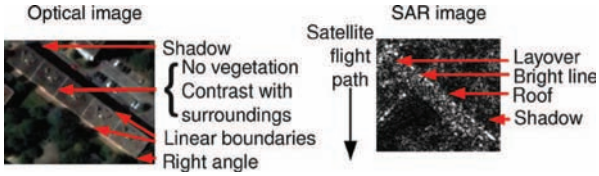


Fig. 2. Characterization of buildings in optical and SAR images.

The goal of the two steps presented in Fig. 1 is to provide a score for each object (coming from the DB or from the segmentation) representing its likelihood of being a building. Consequently, simple cases can be processed automatically, whereas more complex cases may require the intervention of a human operator.

Moreover, a constraint of our algorithm is the absence of learning set to classify building candidates. Therefore, the classification has to be performed with prior knowledge on buildings. However, if a DB is available, the DB verification will be performed with the prior knowledge. Accepted buildings after this step will be used to optimize parameters of the building detection algorithm.

III. Feature Extraction

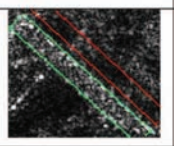
A. Hypothesis Generation

In the DB verification step, we consider objects coming from the DB (for buildings, these objects are polygons). As detailed in Fig. 1, objects are subject to feature computation, feature fusion, and decision. For the detection of new objects (left part of the chain shown in Fig. 1), the three same steps are applied to objects (polygons). To do so, we must generate these objects, i.e., extract object hypotheses from images. In this work, we only extract hypotheses from the optical images. They can also be extracted from SAR images, however, it is more challenging. To generate building hypotheses, we perform a multiscale segmentation of the optical image by using the mean-shift algorithm [23]. This algorithm requires the following three parameters: 1) spatial radius; 2) spectral radius; and 3) minimum region size. We use various spectral radii to obtain segmentations at different scales. The resulting regions are then transformed into polygons that will be used for the detection of new objects. Note that if a DB is available, it was checked in the previous step. Consequently, building hypotheses are regions coming from the segmentation of the optical image that do not intersect polygons kept in the DB.

B. Feature Computation

The goal of feature computation is to find clues in images about the presence of buildings. As the proposed approach must be generic, we have to find features common to most kinds of buildings. As represented in Fig. 2, in optical images, most buildings contrast with their surrounding, cast a shadow (as a building is higher than its surrounding), contain no vegetation, and have linear walls (this feature characterizes man-made structures). In SAR images, some buildings present a contrast between the layover and the shadow area [24]. However, these characteristics are not observed for all buildings. Indeed, if the Sun is at its zenith, elevated objects do not cast a shadow. Moreover, small houses do not have long enough walls to be

TABLE I
FEATURE EXTRACTION

Sensors	Optical				SAR
	Panchromatic		Multispectral		
Properties	Higher than surroundings	Linear structure	Contrasts with surroundings	No vegetation	SAR contrast
Primitives	Shadow mask	Line Segment [27]	Edges [23] and distance to edge measure [29]	Vegetation mask(NDVI [28] thresholding)	Ratio layover/shadow
Examples					

detected as meaningful segments. A fusion process detailed in the next part is used to combine all available information extracted from images. The result of this fusion is used to make a decision on the relevance to include an object in the DB.

Algorithms used to extract primitives associated to this prior information are provided in Table I. Once primitives have been extracted from input images, appropriate features are computed in the vicinity of each building hypothesis providing a score associated to the presence of buildings. These features are listed as follows.

1) *Shadow:* This feature requires to know the direction of the Sun as a prior information on the optical image. This information is used to determine which building hypothesis walls are oriented toward the Sun. A shadow mask is also required. In this work, we obtain the shadow mask by using an empirical thresholding of the optical image. However, automatic methods such as the ones developed in [25] and [26] could also be used for that purpose. We consider building hypothesis wall pixels, which are pixels of the optical image lying on the edges of input polygons. Around each building hypothesis wall not oriented toward the Sun, we define a buffer region (width of several pixels depending on the image resolution) and compute the percentage of wall pixels that contain a shadow pixel in their neighborhood.
2) *Line segments:* Segments are extracted from the optical image using the line segment detector [27]. For each building hypothesis wall (each line of input polygon), we consider extracted segments that are in their neighborhood and parallel to the wall (for our tests we allowed a tolerance of 10°). We compute the percentage of building hypothesis wall pixels containing an extracted segment parallel to the wall in their neighborhood.
3) *Edges:* We compute the contrast between the building hypothesis and its neighborhood. The resulting score is the mean distance between building hypothesis borders and optical image nearest edges (extracted using mean shift [23]).
4) *No vegetation:* This feature requires a multispectral image. We determine the percentage of not vegetated pixels located inside the building hypothesis (a vegetation mask is obtained thanks to a thresholding of the normalized difference vegetation index [28] of the multispectral image).
5) *SAR contrast:* We define buffer regions around walls oriented toward the sensor (layover region) and behind opposite walls (shadow region). For each building hypothesis, we compute the ratio of means: score $= \log(m(\text{layover})/m(\text{shadow}))$.

These five features are examples that have been implemented to provide data to the fusion step, which is the crucial point of this work. Moreover, once the strategy of feature adding has been set up, new features can be easily integrated to the proposed generic processing chain. Note that according to the scenario, some features may not be computable. For instance, if only one optical panchromatic image is available, only three features will be computed (shadow, line segments, and edges). The score indicating the likelihood of being a building is computed from a fusion procedure described in Section IV.

IV. FEATURE FUSION

A. Fusion Framework

As explained in Section III, several features are extracted from images. Each feature brings evidence on the presence of a building. To benefit from all information brought by features, we need to combine scores coming from each feature. The goal of the fusion is to exploit redundancy and to reduce uncertainty. Evidential reasoning can be based on the following three frameworks: 1) Bayesian probability theory [30]; 2) Dempster–Shafer theory of evidence [31]; and 3) possibility theory [32]. The Bayesian probability theory is a classical method for data fusion that is based on a well-developed decision-making theory. However, it requires a considerable amount of prior knowledge and cannot easily model imprecise, incomplete, and not totally reliable information. The Dempster–Shafer theory of evidence is a generalization of probability theory that allows us to capture the imprecise nature of evidence. The resulting decision is not very well defined since degrees of likelihood are measured by probability intervals instead of probabilities for the Bayesian framework. The possibility theory, based on fuzzy set theory [33], is also adapted to uncertain and imprecise information. It might be used as well in our application. However, with the possibility theory, several combination rules are possible, and the choice between these rules is not straightforward. The Dempster–Shafer evidence theory has a clearer and more rigorous foundation. Moreover, it provides interesting byproducts, such as conflict between sources and ignorance (as the confidence is expressed through intervals). These information

can be used to analyze complex cases that do not fit the established model. A human operator might want to focus on those complex cases, while cases with a great confidence value might be processed automatically. In our application, features bring pieces of evidence on the probability of being a building. However, most of them do not discriminate buildings alone (for instance, absence of vegetation can be an evidence of presence of buildings, as well as roads). The Dempster–Shafer theory of evidence appears as the best adapted framework to represent and manage imprecision of features, as well as allow the easy integration of new features in the chain.

B. Evidence Theory

In the Dempster–Shafer framework, evidence is assigned to elements of the set of all possible propositions called *frame of discernment*, often denoted by Θ. The power set $\mathbb{P}(\Theta)$ is the set of all possible subsets of Θ. Subsets of Θ are called *propositions*. The quantity of evidence that a source assigns to a proposition is represented by a *mass function* (MS: also called *basic probability assignment*). An MF m satisfies the following properties:

$$m:\ \mathbb{P}(\Theta) \rightarrow [0,1],\ \sum_{A_i \subseteq \Theta} m(A_i) = 1, m(\emptyset) = 0. \quad (1)$$

Subsets that are assigned a mass by a source are called *focal sets* of the source.

The uncertainty corresponds to the set Θ. Considering the focal set model built according to feature imprecision, masses assigned to sets included in (resp. containing) the building set will constitute the belief (resp. the plausibility) of building hypothesis. After defining focal sets for each feature, the fusion of information is performed, thanks to the Dempster–Shafer orthogonal rule. The mass of a proposition P, resulting from the combination of two sources 1 and 2 is expressed as follows:

$$m_{12}(P) = m_1 \oplus m_2(P) = \frac{1}{1-\kappa} \sum_{A \cap B = P} m_1(A) m_2(B) \quad (2)$$

with $\kappa = \sum_{A \cap B = \emptyset} m_1(A) m_2(B)$.

Information about each proposition is represented by an interval, bounded by two values, namely, 1) the belief and 2) the plausibility. The belief function contains all evidences attached to subsets of the proposition P, i.e.,

$$\text{Bel}(P) = \sum_{A \subseteq P} m(A). \quad (3)$$

The plausibility function is the sum of all the masses that intersect the set of interest P, i.e.,

$$\text{Pl}(P) = \sum_{A | A \cap P \neq \emptyset} m(A). \quad (4)$$

There are various ways to take a decision in the Dempster–Shafer framework. The main decision rules are the maximum of belief, the maximum of plausibility, and the center of the interval whose boundaries are belief and plausibility. For our tests, we have chosen the tradeoff consisting of taking the mean of belief and plausibility.

C. Imprecision Representation

To take into account the imprecision of our features, we build a model representing relationships between focal sets. To build the model, we consider each feature and the type of object they can discriminate. The approach is given as follows. A ground truth of an image is used to generate samples of various classes in an image. In a dense urban environment, we consider typically the following classes: building, vegetation, road, shadow, water, and heterogeneous regions (regions composed of parts of other classes). Feature values for the various classes are used to build histograms. For instance, histograms for the features *Shadow* and *No vegetation* are represented in Fig. 3. The histogram depicted in Fig. 3(a) shows that elements of classes *Roads*, *Shadows*, and *Heterogeneous objects* have a very low value, while elements of classes *Buildings* and *Vegetation* are more spread over the histogram. Consequently, most buildings and vegetation can be distinguished from the other classes. Moreover, the histogram depicted in Fig. 3(b) shows that the feature *No vegetation* allows one to discriminate between vegetation and buildings. More generally, considering histograms for the five features, the analysis of feature value repartition for each class leads to following properties:

1) Buildings contain no vegetation. However, there are other objects that are not vegetated (like roads).
2) Objects contrast with their surrounding.
3) Most buildings and trees project a shadow.
4) Most objects with linear borders are man-made structures (roads or buildings).
5) Only some buildings present a contrast between the layover and the shadow area.

Based on these remarks, we build the model of relationships between focal sets displayed in Fig. 4. This figure shows relationships of partial or total inclusion between focal sets. Even if most buildings contain no vegetation, a nonvegetated object is not always a building. Consequently, the set *Building* (hatched in Fig. 4) is included in the focal set *No vegetation*. We mentioned previously that the feature *Shadow* can distinguish elevated objects (high vegetation and most buildings) from other objects. Consequently, in the prior model of Fig. 4, the only nonvegetated set included in the *Shadow border* focal set is the *Building* set. Other elements of the *Shadow border* set are vegetated objects. In SAR images, according to their size and their orientation, some buildings present a contrast between layover and shadow area. Consequently, the *SAR contrast* focal set is totally included in the *Building* set. The Dempster–Shafer evidence theory assigns pieces of evidence to each focal set. During the fusion step, the mass of each set is computed according to the Dempster–Shafer orthogonal rule. Finally, the decision is taken according to the value of belief and plausibility of the building set (hatched in Fig. 4). This model gathers prior information on feature imprecision. Data fitting will be represented by MFs detailed in the next part.

Note that the Bayesian probability theory would be more complicated to apply. Indeed, it would require to define the conditional probabilities of each subset for the fusion (for instance, as the shadow feature cannot discriminate between buildings and trees, $P(\text{Building}|\text{Cast shadow})$ and $P(\text{Tree}|\text{Cast shadow})$ should be known, which requires a lot of prior knowledge).

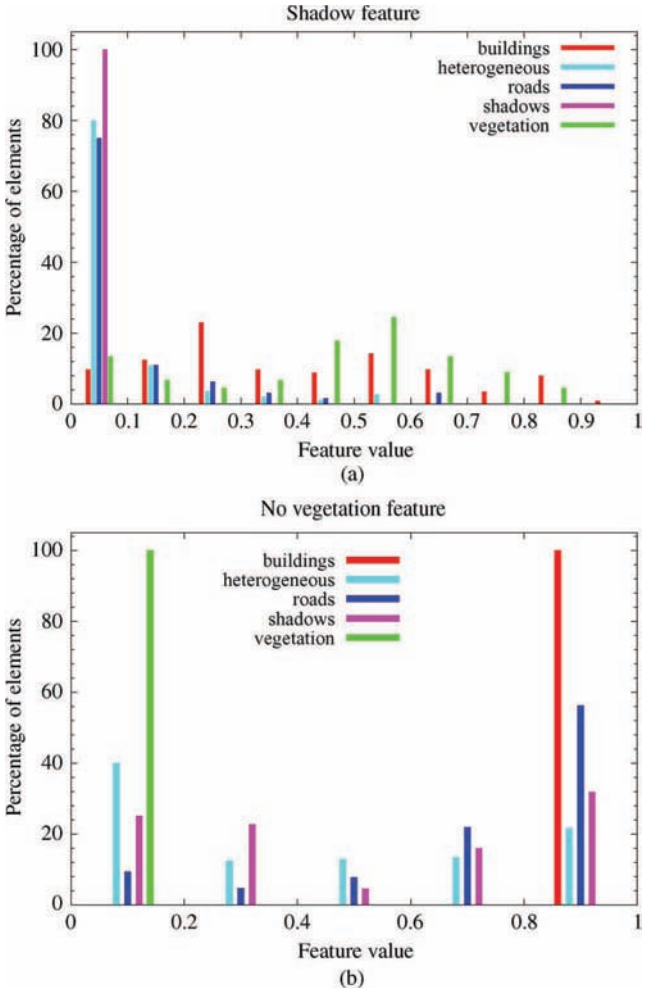


Fig. 3. Examples of histograms for features (a) *Shadow* and (b) *No vegetation*.

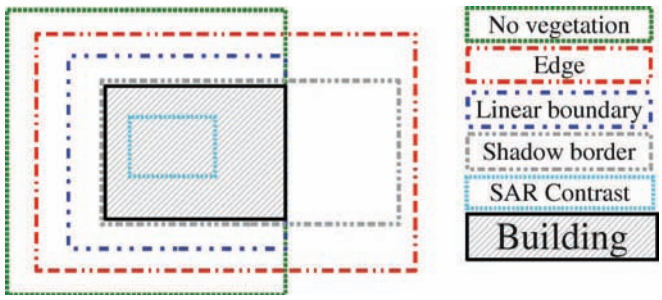


Fig. 4. Model of focal sets.

D. Representation of Uncertainty

MFs have to be defined to represent uncertainty. MFs determine the quantity of evidence brought by a source to each set defined in the model represented in Fig. 4. Each feature that we have implemented brings information to the following three sets: 1) the corresponding focal set; 2) its complementary; and 3) its uncertainty. To determine the quantity of evidence brought to each of these sets, we use trapezoidal MFs that have shown simplicity and efficiency in a similar application [34]. The trapezoidal functions used in our study are represented in Fig. 5. For instance, an object with a high score for the feature *No vegetation* will lead to a high mass for the set *No vegetation* and a low mass for the uncertainty. Conversely, if the object is composed of vegetation pixels in almost half of its surface, the uncertainty will be high for this feature. MFs depend on four parameters (as represented in Fig. 5) gathered in $\theta_j = (a_j, b_j, c_j, d_j)$, where j varies from 1 to 5 and corresponds to each feature. At the beginning of the chain, MF parameters have to be set up. It can be achieved empirically, or optimized, thanks to a ground truth. After the DB verification (if a DB is present), a learning set of verified buildings is available. Section IV-E addresses the problem of optimizing the vector $\theta = (\theta_1, \ldots, \theta_5)$, thanks to a learning set. In our tests, we have used true buildings coming from a ground truth, and false buildings (corresponding to other objects like shadows, roads, trees, and heterogeneous objects). Then, MFs have been optimized using the method detailed in Section IV-E. Note that once MFs have been optimized with an image data set, they can be used to process other images acquired in the same conditions (image type, resolution, environment, illumination, ...).

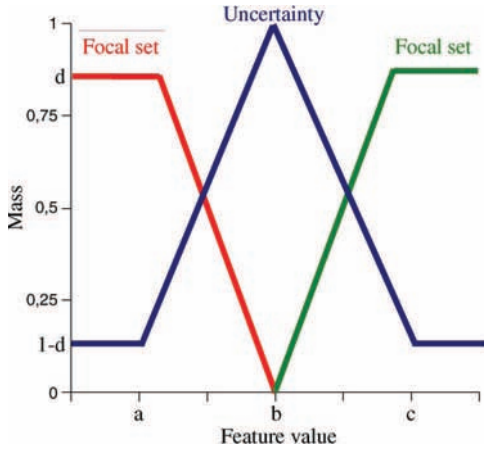


Fig. 5. Model for mass functions.

E. Parameter Optimization

For each feature (shadow, line segment, edge, no vegetation, and SAR contrast), three MFs are needed (for the focal set, its complementary and uncertainty), which lead to four parameters to optimize per feature (so 20 parameters in total). As described in the processing chain shown in Fig. 1, parameter optimization is conducted in the case of an available DB. After verifying the DB, input objects are divided into the following two classes: 1) buildings kept in the DB and 2) buildings removed from the DB. We use those samples to optimize the 20 parameters of the MF model, which is used to perform the detection of new buildings. The function that is minimized to estimate θ is defined as

$$F(\theta) = p \sum_{i=1}^{n_0} \left(1 - \frac{\mathrm{Bel}(B_i) + \mathrm{Pl}(B_i)}{2}\right)^2 + (1-p) \sum_{i=n_0+1}^{N} \left(\frac{\mathrm{Bel}(B_i) + \mathrm{Pl}(B_i)}{2}\right)^2$$

where $B_1, \ldots, B_N$ represent the *Building* focal set of each object. We consider that objects numbered from 1 to n_0 are kept

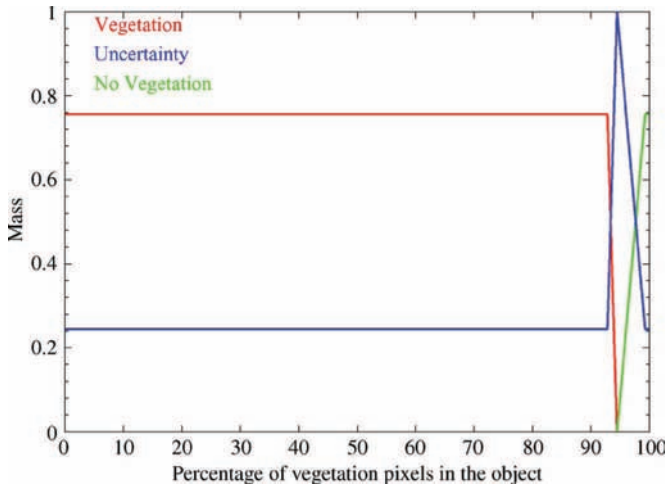


Fig. 6. Optimized mass functions for the feature *No vegetation*.

in the DB, while those numbered from $n_0 + 1$ to N are removed from the DB. Minimizing $F(\theta)$ consists of maximizing the mean of belief and plausibility for accepted buildings and minimizing this mean for rejected buildings, with p weighting the two terms appearing in F (p is set to 0.5 in this paper reflecting the absence of knowledge for this parameter). The optimization can be achieved using a numerical optimization procedure. Results presented in this paper have been obtained, thanks to the Nelder–Mead method [35].

MFs optimized for the feature *No vegetation* are provided in Fig. 6. The four parameters for this feature are estimated using real buildings (coming from an up-to-date DB) and false buildings (corresponding to other objects). Parameters are initialized using prior information: We consider that the uncertainty is maximum when half of object pixels are vegetation pixels ($b = 50$). Moreover, we have chosen $a = 0$ and $c = 100$ and a reliability (parameter d) equal to 0.8 (as we have no information about these parameters). The result of the optimization proves that most of buildings contain no vegetation. Indeed, if an object contains at least 5.5% of vegetated pixels ($b = 94.5$), no evidence will be brought to the *No vegetation* focal set. The reliability of this feature has a value of 0.75. Consequently, a highly vegetated building will present an uncertainty of 0.25, keeping a possibility to be considered as a building if it has other characteristics of a building (like a contrast in SAR images, line parallel to its wall, . . .).

Note that if some wrong buildings are kept in the DB or if true buildings are rejected, the parameter estimation will be impacted. According to the application, solutions are possible to overcome this issue. For instance, a human operator can be included in the chain after the DB verification step to validate results. He will focus on uncertain and complex cases. Consequently, the learning step will be performed with a correct learning set. If the chain is used in a fully automatic way, errors after the DB verification will reduce the quality of the object detection step.

V. Simulation Results

Experiments have been performed to evaluate both the DB verification and the detection of new buildings. Two data sets are available. The first one has been acquired over Toulouse, France, in an urban environment. More precisely, we used a Pleiades-simulated image (coming from the airborne sensor PELICAN downsampled to 70 cm and to 2.5 m) and a TerraSAR-X satellite image at 1 m resolution associated to the same area. The test area contains 111 buildings, i.e., 70 982 pixels in the 70-cm-resolution optical image (whose size is 787 × 888 pixels). The second data set has been acquired over Haiti. It consists of a 60-cm QuickBird image and a TerraSAR-X image at 1 m resolution. This test area contains 100 buildings, i.e., 175 824 pixels in the 60-cm-resolution optical image (whose size is 1100 × 1332 pixels). Building DBs contain vector data representing the 2-D coordinates of building outlines.

Results are evaluated both at object and pixel levels. At the object level, a reference building is considered as detected if more than 50% of its surface has been detected. A detected object is considered as a false alarm if more than 50% of its surface does not correspond to a building in the ground truth. Note that this threshold of 50% is a balanced value used in our tests. However, a stricter criterion might be used, for instance, by considering a building as detected if at least 90% of its surface has been detected, and an object as a false alarm if more than 10% of its surface does not correspond to a building. Such a strict criterion would tolerate less imprecision in the detected building delineation. In our tests, it would not impact the DB verification results because we use a ground truth to provide true buildings. A correctly accepted building is a true positive (TP). A correctly rejected object is a true negative (TN). A wrongly rejected building is a false negative (FN). A wrongly accepted object is a false positive (FP). Considering the number of TP, TN, FN, and FP for each case, we compute precision and recall [36]. The precision can be seen as a measure of correctness (corresponds to the probability that an accepted object is a true building), whereas the recall is a measure of completeness (corresponds to the probability that a reference building is accepted). The closer those coefficients are to 1, the better the result will be. The F-measure corresponds to the harmonic mean of precision and recall. The pixel level evaluation is performed using the false alarm rate (FAR) and the detection rate (DR). It is a classical evaluation method in this type of work. However, this evaluation is not suitable for the object level evaluation of new building detection. Indeed, FAR depends on the number of TNs (false buildings correctly rejected). In the building detection step, the number of false buildings is huge (they are due to oversegmentation). Consequently, FAR is always very low and, thus, is not interesting to evaluate results. The advantage of recall/precision evaluation is its independence to the number of TNs.

For the first data set, acquired over Toulouse, results are presented in the following seven cases, which gather possible scenarios of our processing chain:

- Case 1: a multispectral 70-cm-resolution image and a 1-m-resolution TerraSAR-X image;
- Case 2: a multispectral 70-cm-resolution image;
- Case 3: a panchromatic 70-cm-resolution image and a 1-m-resolution TerraSAR-X image;
- Case 4: a panchromatic 70-cm-resolution image;
- Case 5: a multispectral 2.5-m-resolution image and a 1-m-resolution TerraSAR-X image;
- Case 6: a multispectral 2.5-m-resolution image;
- Case 7: a 1-m-resolution TerraSAR-X image.

TABLE II
DB VERIFICATION RESULTS USING TOULOUSE DATA SET

Scenario	Object level evaluation							Pixel level evaluation	
	TP	TN	FN	FP	Precision	Recall	F-measure	FAR	DR
Case 1	108	145	3	3	0.973	0.973	0.973	0.00455	0.938
Case 2	108	145	3	3	0.973	0.973	0.973	0.00420	0.920
Case 3	103	140	8	8	0.928	0.928	0.928	0.00790	0.920
Case 4	99	140	12	8	0.925	0.892	0.908	0.00823	0.900
Case 5	101	140	10	8	0.927	0.910	0.918	0.0127	0.841
Case 6	97	140	14	8	0.924	0.874	0.898	0.0123	0.826
Case 7	82	129	29	19	0.812	0.739	0.774	0.0224	0.811

For the second data set, acquired over Haiti, scenarios are given as follows:

Case 1: a multispectral 60-cm-resolution QuickBird image and a 1-m-resolution TerraSAR-X image;
Case 2: a multispectral 60-cm-resolution QuickBird image;
Case 3: a panchromatic 60-cm-resolution QuickBird image and a 1-m-resolution TerraSAR-X image;
Case 4: a panchromatic 60-cm-resolution QuickBird image;
Case 5: a 1-m-resolution TerraSAR-X image.

The proposed decision is taken thanks to a threshold T that can vary in the interval [0, 1]. An object is accepted if

$$\frac{\mathrm{Pl}(B) + \mathrm{Bel}(B)}{2} \geq T. \tag{5}$$

This threshold is empirical and depends on the application where a low FAR or a high DR may be preferred. It can be adjusted by a human operator accordingly, or kept to a default value in a fully automatic use of the processing chain.

A. DB Verification

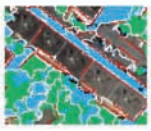
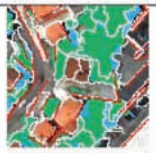
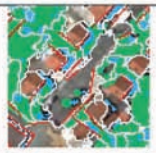
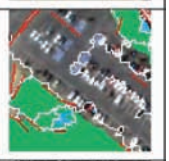
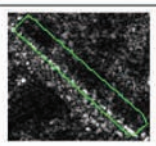
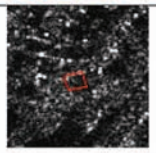
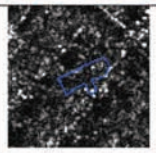
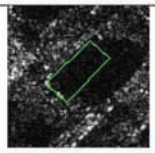
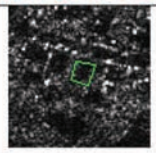
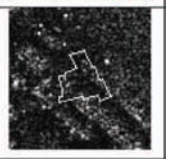
To evaluate the DB verification with the data set acquired over Toulouse, we have created a DB composed of true building outlines coming from a ground truth (111 buildings) and of false buildings (polygons) created manually (148 objects). The evaluation consists of checking if the processing chain is able to keep true buildings and remove false buildings. Results are presented in Table II for the seven cases. Our results prove that the DB verification step can be performed efficiently in the first six cases. However, the problem is more challenging when only a SAR image is available. The analysis of cases 5 and 6 highlights the interest of a 1-m-resolution SAR image when the optical image has a resolution of 2.5 m. However, when the optical image has an HR (70 cm), cases 1 and 2 show that the presence of a 1-m-resolution SAR image does not improve results. Fig. 7 illustrates the DB verification procedure for the first case. The threshold T mentioned in (5) was tuned empirically to have a balanced result, with three FPs and three FNs. This result was obtained with a threshold value of 0.25.

To obtain those results, MFs have been determined, thanks to a parameter optimization based on samples coming from a ground truth. Some results associated to the DB verification for case 1 are presented in Table III. This table shows interesting objects [referred to as case (a)–(f)] extracted from Fig. 7 and the corresponding masses brought by each feature. The third and fifth rows of this table correspond to the hypothesis of building outline projected, respectively, on the optical and SAR images. The fourth row corresponds to primitives extracted from the optical image with the following colors: edges in white; line segments in red; vegetation mask in green; and shadow mask in blue. The considered cases have the following characteristics.

Fig. 7. Result of DB verification with the Toulouse data set for case 1, with the following colors: TP (in green), 108 buildings correctly accepted; FN (in red), three buildings wrongly rejected; FP (in blue), three objects wrongly accepted; and TN (in white), 145 objects correctly rejected.

1) Case (a) corresponds to a true building correctly kept in the DB (like 107 other buildings). It is a large and high building that contrasts with its neighborhood, projects a large shadow, with linear boundaries, contains no vegetation and presents a layover and a shadow area in the SAR image. Masses assigned to each focal set highlight the presence of all characteristics for this building. Consequently, the score corresponding to the mean of belief and plausibility for the set *Building* is high (0.87).
2) Case (b) corresponds to a real building that has been removed from the DB (like two other buildings). MFs show that it is caused by the absence of cast shadow (because of high vegetation surrounding the house), the absence of noticeable contrast in SAR image, and the absence of linear boundaries.
3) Case (c) corresponds to a false building accepted in the DB (like two other false buildings). The polygon is located in a heterogeneous region, close to edges and linear boundaries due to the road. Note that the belief value of the *Building* set is the same as that for the previous case. However, this case is accepted, thanks to

TABLE III
SIMULATION RESULTS

Cases		(a)	(b)	(c)	(d)	(e)	(f)
Status		true positive	false negative	false positive	true positive	true positive	true negative
Optical image							
Extracted primitives							
SAR image							
Shadow	m(Shadow)	0.74	0	0	0.67	0	0
	m($\overline{\text{Shadow}}$)	0	0.39	0.41	0	0.30	0.48
	m(Θ)	0.26	0.61	0.59	0.33	0.70	0.52
Linear boundary	m(Linear bnd.)	0.81	0	0.11	0.81	0.81	0
	m($\overline{\text{Linear bnd.}}$)	0	0.38	0	0	0	0.12
	m(Θ)	0.19	0.62	0.89	0.19	0.19	0.88
Edges	m(Edges)	0.25	0.46	0	0	0.51	0
	m($\overline{\text{Edges}}$)	0	0	0.03	0.81	0	0.71
	m(Θ)	0.75	0.54	0.97	0.19	0.49	0.29
No vegetation	m(No veg.)	0.71	0.55	0	0.73	0.70	0.70
	m($\overline{\text{No veg.}}$)	0	0	0.09	0	0	0
	m(Θ)	0.29	0.45	0.91	0.27	0.30	0.30
SAR contrast	m(SAR cont.)	0.12	0	0	0.61	0	0
	m($\overline{\text{SAR cont.}}$)	0	0.11	0.82	0	0.21	0.36
	m(Θ)	0.88	0.89	0.18	0.39	0.79	0.64
Conflict		0	0	0	0.13	0	0
Bel(building)		0.73	0	0	0.78	0	0
Pl(building)		1	0.38	0.53	0.90	0.70	0.13
Decision		0.87	0.19	0.27	0.84	0.35	0.067

its high plausibility value, due to the high uncertainty of most features.

4) Case (d) corresponds to a true building that has been accepted with a high value of conflict. As explained previously, the conflict is an interesting byproduct of the Dempster–Shafer evidence theory. Conflict denotes an error in the model, or a problem of feature reliability. In this case, the building is high and projects a large shadow. However, in the optical image, its radiometry is very close to the radiometry of the adjacent road. Consequently, edges between the building and the road are not detected by the edge detector. The polygon is considered as located far from edges, and a conflict appears between the features "shadow" and "edges." For this situation, a human operator should be alerted to process himself complex cases that do not correspond to the model. Indeed, risks to have a wrong automatic decision in those cases is high.
5) Case (e) corresponds to a small house, correctly kept in the DB. As this house is very low, its shadow is not visible. Moreover, in the SAR image, dimensions of the house are too small to present a significant contrast between the layover and the shadow area. However, this house contrasts with its neighborhood, linear boundaries are present, and it contains no vegetation. Therefore, even if the belief of the *Building* set is equal to 0, the plausibility is high enough to be kept in the DB.
6) Case (f) corresponds to a false building correctly removed from the DB (like 144 other false buildings). The building hypothesis has been positioned in a parking lot. Therefore, it contains no vegetation, the outline is far from the edges, it does not project a shadow, it has no linear boundaries, and it does not present a contrast in the SAR image. Thus, the plausibility of the *Building* set is low, and the object has been rejected.

The evaluation of the DB verification with the Haiti data set is presented in Table IV. The DB was composed of 100 true buildings and 50 false buildings created manually. Results presented in Table IV confirm the ability of the processing chain to perform the DB verification in various environments. Fig. 8 represents accepted and rejected building hypotheses

TABLE IV
DB VERIFICATION RESULTS USING HAITI DATA SET

Scenario	Object level evaluation							Pixel level evaluation	
	TP	TN	FN	FP	Precision	Recall	F-measure	FAR	DR
Case 1	97	48	3	2	0.980	0.970	0.975	0.00738	0.967
Case 2	97	47	3	3	0.970	0.970	0.970	0.00762	0.967
Case 3	93	46	7	4	0.959	0.930	0.944	0.00668	0.954
Case 4	93	45	7	5	0.949	0.930	0.939	0.00923	0.954
Case 5	79	34	21	16	0.832	0.790	0.810	0.0220	0.824

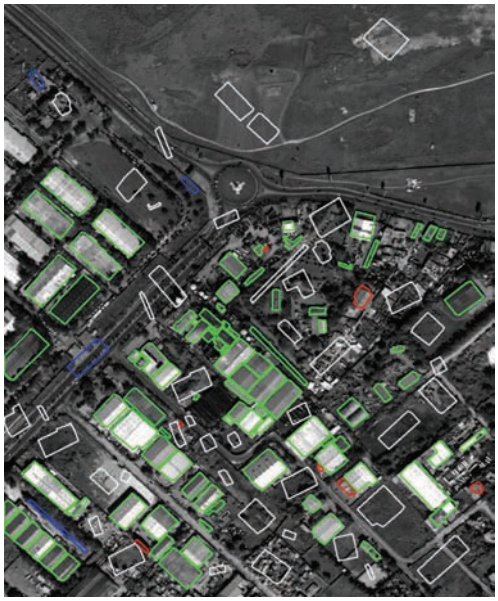


Fig. 8. Result of DB verification with the Haiti data set for case 3, with the following colors: TP (in green), 93 buildings correctly accepted; FN (in red): seven buildings wrongly rejected; FP (in blue), four objects wrongly accepted; and TN (in white), 46 objects correctly rejected.

for the third case (a panchromatic 60-cm-resolution QuickBird image and a 1-m-resolution TerraSAR-X image). This figure shows that most FN buildings are very small. Their primitives are hard to detect, even in a 60-cm-resolution image. FN objects are located near linear edges. As in this scenario (case 3), the multispectral information is not available, and the presence of vegetation cannot be used to reject building hypotheses located over a vegetated area.

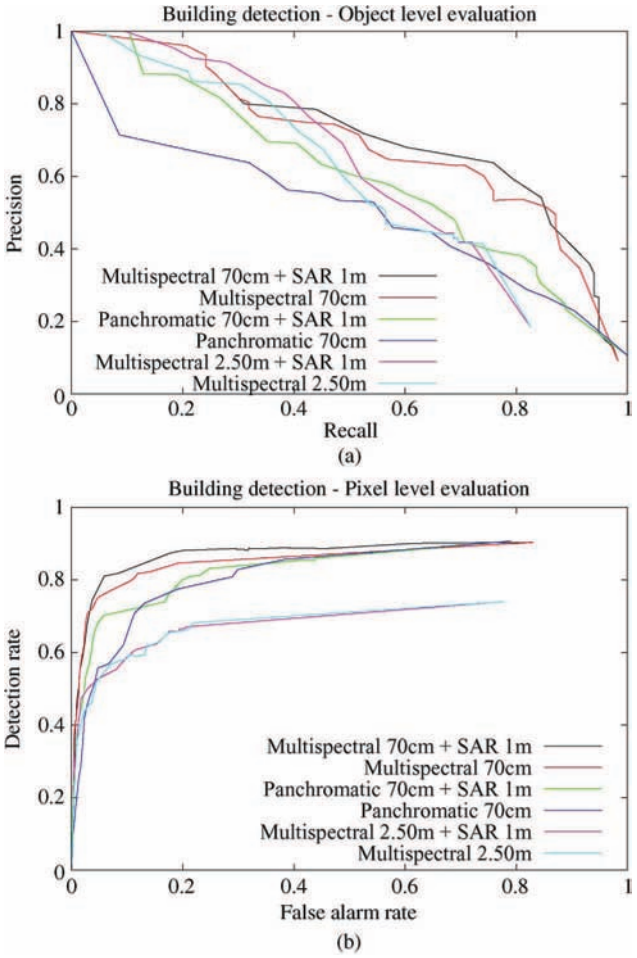


Fig. 9. Building detection: evaluation (a) at object level with precision–recall curves and (b) at pixel level with receiver operating characteristic curves.

B. Building Detection

The previously checked DB was finally used for parameter optimization. To evaluate the detection of new buildings, we no longer use this DB. Indeed, in an operational use of this chain, we could try to detect buildings that are not in the DB. However, this number of buildings is very small after the DB verification. To have more representative results, we have considered an empty DB. In this case, the ideal goal of the chain is to detect all buildings in the images, i.e., to create the DB. In this step, we consider polygons provided by the multiscale segmentation of the optical image. These polygons are processed similarly to DB objects, as detailed in the previous part.

Results are presented in Fig. 9 for the first six cases related to the Toulouse data set described previously. The case where only a SAR image is available has not been considered here because there was no building extraction method from SAR images available in the free image processing library Orfeo Toolbox. The object level evaluation is represented with precision–recall curves in Fig. 9(a). The pixel level evaluation is represented in Fig. 9(b). These curves are obtained by changing the value of the threshold T defined in (5). Curves shown in Fig. 9 confirm that the availability of a SAR image slightly improves results. Moreover, Fig. 9(a) shows that, contrary to the DB verification, building detection is performed better with a 70-cm-resolution panchromatic image than with a 2.5-m-resolution multispectral image. In the building detection step, the spectral information seems more important than the spatial resolution. This is due to the segmentation step, which is used to generate building hypotheses. A good delineation of buildings is hard to achieve with the segmentation of a panchromatic image. Results show

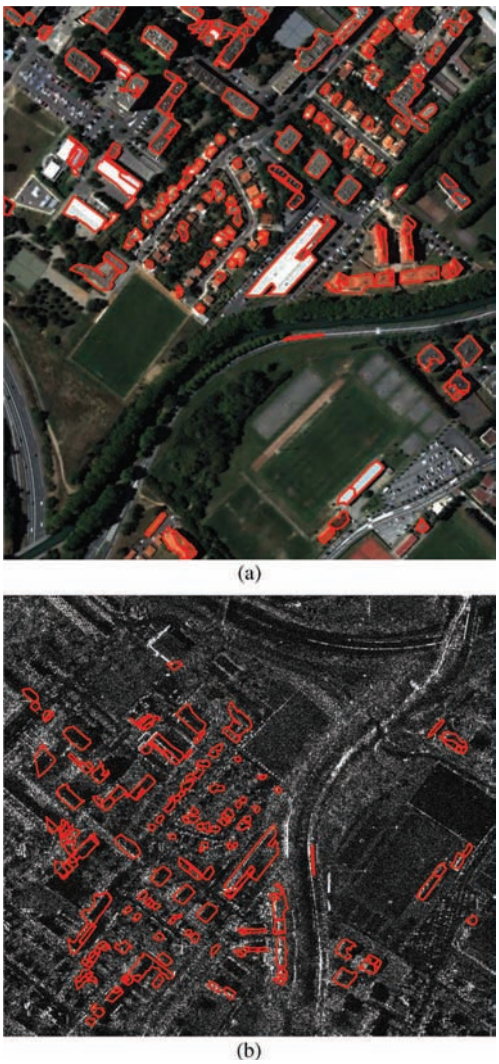


Fig. 10. Building detection without DB: building outlines projected (a) on the optical image and (b) on the SAR image.

that for the first case, at pixel level, almost 70% of building pixels are detected for only 2% of false alarm. This result corresponds to a precision of 72% and a recall of 62%. Detected buildings for this result are represented in Fig. 10. Results projected on the optical image [Fig. 10(a)] and on the SAR image [Fig. 10(b)] highlight that small houses are hard to detect. When buildings are correctly segmented in the optical image, results are very close to those obtained in the DB verification step. It corresponds to buildings that present a high contrast with their neighborhood. On the contrary, Fig. 10(a) shows that some large buildings whose radiometry is close to the adjacent road are not correctly detected. This is due to the problem of segmentation. Indeed, at a coarse scale, those buildings are merged with the adjacent road. However, at a fine scale, they are divided into small regions, which do not verify characteristics of a building.

VI. Conclusion

This paper has described a generic processing chain to update/create a cartographic DB with SAR and optical input images. Our results showed that the chain can process images in a large range of resolution (tested at 0.6 and 2.5 m resolution for the optical image and at 1m for the SAR image). The chosen fusion framework was well adapted to the representation of feature imprecision. New features can be easily included in the proposed chain to improve the building detection. A score for each building hypothesis based on belief and plausibility gave a confidence value for each DB element. It also allowed a human expert to focus only on complex cases while simple elements where processed automatically. Future work will consist of analyzing these complex cases to extract new features able to discriminate these elements. For instance, new features could characterize the shape of segmented regions in the optical image. This will be useful to discriminate buildings and vegetation if just a panchromatic image is available. A similar processing chain is currently under investigation to perform the update of road DBs. Finally, it is interesting to mention that the proposed building processing chain was implemented using CNES ORFEO Toolbox free software http://www.orfeo-toolbox.org.

References

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

[2] A. Katartzis and H. Sahli, "A stochastic framework for the identification of building rooftops using a single remote sensing image," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 259–271, Jan. 2008.

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

[4] Z. J. Liu, J. Wang, and W. P. Liu, "Building extraction from high resolution imagery based on multi-scale object oriented classification and probabilistic Hough transform," in *Proc. IEEE IGARSS*, Seoul, Korea, Jul. 2005, pp. 2250–2253.

[5] W. Liu and V. Prinet, "Building detection from high-resolution satellite image using probability model," in *Proc. IEEE IGARSS*, Seoul, Korea, Jul. 2005, pp. 3888–3891.

[6] S. Muller and D. W. Zaum, "Robust building detection in aerial images," in *Proc. IAPRS*, Vienna, Austria, Aug. 2005, pp. 143–148.

[7] B. Sirmacek and C. Unsalan, "Building detection from aerial images using invariant color features and shadow information," in *Proc. 23rd ISCIS*, Oct. 2008, pp. 1–5.

[8] T. Bailloeul, V. Prinet, B. Serra, P. Marthon, P. Chen, and H. Zhang, "Urban building land use change mapping from high resolution satellite imagery, active contours and Hough voting," in *Proc. Ninth ISPMSRS*, Beijing, China, 2005.

[9] O. Aytekin, I. Ulusoy, E. Z. Abacioglu, and E. Gokcay, "Building detection in high resolution remotely sensed images based on morphological operators," in *Proc. Fourth Int. Conf. RAST*, Jun. 2009, pp. 376–379.

[10] S. Lefevre, J. Weber, and D. Sheeren, "Automatic building extraction in VHR images using advanced morphological operators," in *Proc. IEEE/ISPRS Joint Workshop Remote Sens. Data Fusion Over Urban Areas (URBAN)*, Paris, France, Apr. 2007, pp. 1–5.

[11] H. Sportouche, F. Tupin, and L. Denise, "Building extraction and 3D reconstruction in urban areas from high-resolution optical and SAR imagery," in *Proc. Urban Remote Sens. Event*, Shanghai, China, May 2009, pp. 1–11.

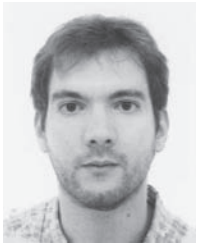
[12] K. Karantzalos and N. Paragios, "Recognition-driven two-dimensional competing priors toward automatic and accurate building detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 133–144, Jan. 2009.

[13] B. Sirmacek and C. Unsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1156–1167, Apr. 2009.

[14] T. Kim and J. P. Muller, "Development of a graph-based approach for building detection," *Image Vis. Comput.*, vol. 17, no. 1, pp. 3–17, Jan. 1999.

[15] M. Quartulli and M. Datcu, "Information extraction from high resolution SAR data for urban scene understanding," in *Proc. Second GRSS/ISPRS Joint Workshop Data Fusion Remote Sens. Over Urban Areas (URBAN)*, 2003, pp. 115–119.

[16] M. Quartulli and M. Datcu, "Stochastic geometrical modeling for built-up area understanding from a single SAR intensity image with meter resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 9, pp. 1996–2003, Sep. 2004.
[17] C. Tison, F. Tupin, and H. Maitre, "Retrieval of building shapes from shadows in high resolution SAR interferometric images," in *Proc. IEEE IGARSS*, Anchorage, AK, Sep. 2004, pp. 1788–1791.
[18] 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.
[19] A. Thiele, E. Cadario, K. Schulz, U. Thoennessen, and U. Soergel, "Building recognition from multi-aspect high-resolution InSAR data in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3583–3593, Nov. 2007.
[20] A. J. Bennett and D. Blacknell, "The extraction of building dimensions from high resolution SAR imagery," in *Proc. Int. Radar Conf.*, Sep. 2003, pp. 182–187.
[21] F. Tupin and M. Roux, "Detection of building outlines based on the fusion of SAR and optical features," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 1, pp. 71–82, Jun. 2003.
[22] V. Amberg, M. Spigai, M. Coulon, and P. Marthon, "Improvement of road extraction in high resolution SAR data by a context-based approach," in *Proc. IEEE IGARSS*, Seoul, South Korea, Jul. 2005, pp. 490–493.
[23] D. Comaniciu and P. Meer, "Mean shift analysis and applications," in *Proc. Seventh IEEE Int. Conf. Comput. Vis.*, Kerkyra, Greece, 1999, pp. 1197–1203.
[24] D. Massonnet and J. C. Souyris, *Imaging With Synthetic Aperture Radar*. Boca Raton, FL: CRC Press, 2008, ch. 3, pp. 165–167.
[25] K. S. Gudmundsson and F. Cagatin, "Complex shadow extraction," in *Proc. SPIE*, San Diego, CA, Aug. 2008, vol. 7072, pp. 707 211.1–707 211.10.
[26] H. T. Guo, Y. Zhang, J. Lu, and G. W. Jin, "Research on the building shadow extraction and elimination method," in *Proc. IAPRS*, Beijing, China, 2008, pp. 569–574.
[27] R. Grompone Von Gioi, J. Jakubowicz, J. M. Morel, and G. Randall, "On straight line segment detection," *J. Math. Imag. Vis.*, vol. 32, no. 3, pp. 313–347, Nov. 2008.
[28] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, May 1979.
[29] P. E. Danielsson, "Euclidean distance mapping," *Comput. Graphics Image Process.*, vol. 14, no. 3, pp. 227–248, Nov. 1980.
[30] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. Hoboken, NJ: Wiley, 1973, pp. 98–105.
[31] P. Smets, "What is Dempster–Shafer's model?" in *Advances in the Dempster–Shafer Theory of Evidence*. New York: Wiley, 1994, pp. 5–34.
[32] D. Dubois and H. Prade, "Possibility theory, probability theory and multiple-valued logics: A clarification," *Ann. Math. Artif. Intell.*, vol. 32, no. 1–4, pp. 35–66, Aug. 2001.
[33] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets Syst.*, vol. 100, no. Suppl. 1, pp. 9–34, 1999.
[34] F. Tupin, I. Bloch, and H. Maitre, "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.
[35] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, Jan. 1965.
[36] J. Davis and M. Goadrich, "The relationship between precision–recall and ROC curves," in *Proc. 23rd ICML*, New York, 2006, pp. 233–240.



Vincent Poulain was born in Dieppe, France, in 1983. He received the Eng. degree in electrical engineering and signal processing from the Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique et d'Hydraulique de Toulouse (ENSEEIHT), Toulouse, France, in 2007 and the Ph.D. degree in image processing from the National Polytechnic Institute of Toulouse in 2010.

In 2010, he joined the Thales Group, where he is currently working in the field of remote sensing image processing.

Jordi Inglada (M'98) received the Eng. degree in telecommunications engineering from both the Universitat Politècnica de Catalunya, Barcelona, Spain, and the École Nationale Supérieure des Télécommunications de Bretagne, Brest, France, in 1997 and the Ph.D. degree in signal processing and telecommunications from Université de Rennes 1, Rennes, France, in 2000.

He is currently with the French Space Agency, Toulouse, France, working in the field of remote sensing image processing at the CESBIO laboratory. He is in charge of the development of image processing algorithms for the operational exploitation of earth observation images, mainly in the field of multitemporal image analysis for land use and cover change.

Dr. Inglada is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Marc Spigai was born in Marseille, France, 1965. He received the Eng. degree in signal and image processing from the École Nationale Supérieure des Télécommunications de Bretagne (Telecom Bretagne), Brest, France, in 1989.

From 1990 to 1998, he works on signal and data processing at Thales System Airborne, Elancourt, France. From 1998 to 2000, he was Head of a research team on fusion of radar, vision, and navigation for driving assistance at Renault, Guyancourt, France. He joined Thales Alenia Space, Toulouse, France, in 2000, where he currently works on SAR signal and image processing at the Observation Systems and Radars Unit.

Jean-Yves Tourneret (SM'08) received the Eng. degree in electrical engineering from the Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique et d'Hydraulique de Toulouse (ENSEEIHT), Toulouse, France, in 1989 and the Ph.D. degree from the National Polytechnic Institute of Toulouse in 1992.

He is currently a Professor with the University of Toulouse and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal processing with a particular interest to MCMC methods.

Dr. Tourneret was the Program Chair of the European Conference on Signal Processing (EUSIPCO). He was also a member of the Organizing Committee for the 2006 International Conference on Acoustics, Speech, and Signal Processing, which was held in Toulouse. He has been a member of different technical committees, including the Signal Processing Theory and Methods committee of the IEEE Signal Processing Society (2001–2007, 2010–present). He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Philippe Marthon received the Eng. and Ph.D. degrees in computer science from the Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique et d'Hydraulique de Toulouse (ENSEEIHT), Toulouse, France, in 1975 and 1978, respectively, and the Doctorat d'Etat degree from the Institut National Polytechnique de Toulouse (INPT), Toulouse, in 1987.

He is currently a Professor at the ENSEEIHT and a Research Scientist at the Institut de Recherche en Informatique de Toulouse (IRIT). He is the author of over 80 scientific articles published in international journals and conference proceedings. His research interests include systemics, operations research, remote sensing image processing, and computer vision.