Information Fusion for Scene Understanding From Interferometric SAR Data in Urban Environments

M. Quartulli and M. Datcu

***Abstract*—We present a framework for scene understanding from interferometric synthetic aperture radar data that is based on Bayesian machine learning and information extraction and fusion. A generic description of the data in terms of multiple models is automatically generated from the original signals. The obtained feature space is then mapped to user semantics representing urban scene elements in a supervised step. The procedure is applicable at multiple scales. We give examples of urban area classification and building recognition of Shuttle Radar Topography Mission data and of building reconstruction from submetric resolution Intermap data.**

***Index Terms*—Bayesian data fusion, interferometric synthetic aperture radar (InSAR), scene understanding.**

I. Introduction

SCENE UNDERSTANDING (SU) is the procedure of finding the scene that best explains a given dataset

SCENE → DATA direct mode of acquisition
SCENE ← DATA scene understanding.

Examples of SU include building recognition and the generation of land use maps from remote sensing data.

When the reconstruction is carried out from measurements derived from multiple sensors of dissimilar nature, or from sources that are fundamentally different—such as radar images and records of knowledge gathered from humans—the techniques of information fusion are used to derive conclusions and decisions based on a synergy of all the data available.

A limitation of deterministic SU algorithms is that the imperfect knowledge of the parameters of the acquisition—as well as the presence of incertitudes and noise in the available measurements—generates an inherent incompleteness of the data.

To be able to deal with incomplete data, we approach the problem of scene understanding and information fusion in a probabilistic framework.

In particular, we consider the ill-posed nature of the SU problem by adopting regularization strategies that incorporate external sources of information into the available data set.

In this context, the data d and the scene s are considered to be samples drawn from the multidimensional random processes D and S [4], [18].

Manuscript received September 22, 2002; revised February 13, 2003. Part of this work was supported by the European Community project Virtual-Planet and collaboration with the Swiss Federal Institute of Technology ETH Zürich.

The authors are with the IMF-BW Remote Sensing Technology Institute-Image Science, DLR German Aerospace Centre, Oberpfaffenhofen 82234, Wessling, Germany.

Digital Object Identifier 10.1109/TGRS.2003.814630

A treatment of the SU inversion problem that incorporates and composes models for these processes is obtained starting from

$$p(s \mid d)p(d) = p(d \mid s)p(s) \tag{1}$$

a relation linking conditional probabilities that is the well-known Bayes' equation.

The main characteristic of the Bayesian approach is the explicitation of the modeling assumptions made for both the data formation mechanism $S \rightarrow D$ in the likelihood term $p(d \mid s)$ and for the phenomenon under analysis S in the prior $p(s)$. Further important features include the capacity to deal with incomplete data, to learn causal relationships—thus learning the problem domain, and the power to combine knowledge and signal information.

The inversion can be carried out by maximization of the Posterior probability, resulting in the so called maximum *a posteriori* (MAP) estimate of the phenomenon S

$$\hat{s}_{\mathrm{MAP}} = \arg\max_{s}\{p(d \mid s)p(s)\}. \tag{2}$$

The MAP equation in (2) can be considered as a model for a learning device [6]: a description for the scene S is obtained from the data d by learning about their relation with the model S. Probability is used as a measure for the uncertainty in the knowledge of this relation.

The SU procedure described in the remainder of this paper is applied to the understanding of urban scenes starting from interferometric synthetic aperture radar (InSAR) data.

The all-weather, all-time characteristics of the SAR, as well as its peculiar sensitivity to scene characteristics such as dielectric properties and both large and small scale geometry [5], have made it the instrument of choice in a number of remote sensing applications.

A particularly important role has been played in recent years by SAR interferometry, an extension of the radar concept that is made possible by the coherent nature of the signal [3]. The contextual exploitation of data acquired on the same area from a number of slightly different positions allows the measurement of the local distance between the scene elements and the interferometer.

An indication of the geometric and radiometric stability of the scene is obtained by considering the interferometric coherence, an estimate of the amplitude of the normalized cross-correlation between the observations [20].

The processing of SAR and InSAR data for the generation of backscatter intensity, digital elevation model (DEM), and interferometric coherence measurements is a well-established field that has attained the operational stage since a number of years.

The exploitation of this data for the understanding of urban scenes is instead a very active area of research. Many studies on the subject focus on techniques for the regularization of height data for the reconstruction of buildings given the very peculiar phenomenology of SAR in built-up areas [10], [13], [17]. More recently, a class of algorithms was proposed for the extraction of specific targets such as road networks [9], [11], and efforts were devoted to the study of basic signal properties of very high-resolution SAR in urban environments [12]. The problem of building reconstruction from InSAR data was also investigated in the specific case of large industrial structures [7] or as a pre-processing step for further exploitation of optical data [1]. Simulation-based reconstruction approaches [2] were also considered. Very recently, the validation of InSAR-derived building maps is starting to be considered with the definition of algorithms for the preparation of reference data [8].

In this work, we give examples by considering an interferometric dataset acquired in the context of the Shuttle Radar Topography Mission (SRTM) on the city of Baltimore, MD. The mission, occurred February 11–22, 2000, was based on the single-pass acquisition of multiple streams of data from two sets of antennas on board of the Space Shuttle. The 12 TB of raw data acquired are currently being processed into digital elevation maps, providing the first-ever global coherent DEM of the earth between latitudes $-80^{\circ}, +80^{\circ}$ together with radar backscatter and interferometric coherence information. The dataset under study, with a spatial resolution of about 25 m, was acquired at X-band.

A very different dataset, acquired on the outskirts of Munich, Germany, from an Intermap airborne sensor with a spatial resolution of about 0.5 m again at X-band, is used as a second example. We consider the results obtained with this second dataset indicative of the quality of the developed algorithms for their application to the large amounts of data that will be acquired at comparable resolutions by the next generation of satellite-borne SAR systems in the next few years.

This paper proceeds as follows. Section II describes our modeling approach for the integration of the available information; Sections III and IV detail the different processing stages that compose our SU device, while Sections V–VII demonstrate our approach on the test datasets.

II. Hierarchical Modeling for InSAR Data

When models for the scene S and the data D become so complicated that the inversion in terms of (2) becomes untractable, a common solution is a divide-and-conquer approach in which simpler intermediate description levels are introduced in the modeling.

The large amount of data required to perform a full inference while keeping high-resolution detail observable would prohibit the treatment of high-order probability distribution functions. This explosion of computational complexity, known as the "curse of dimensionality," has to be treated by recurring to models described by only local statistics, as is the case in Markov models and of hierarchical Bayesian networks.

Bayesian networks [6] allow us to overcome the curse of dimensionality by encoding probabilistic independence between the elements of the inference problem. They can be mapped to graphs whose nodes represent the random variables under study, while dependencies are encoded in their structure. The learning procedure progressively updates the network thereby obtaining a description of the statistical link between the variables based on all the information available [Fig. 1(a)]. Once the graph structure is set up, it can be used to perform efficient inference by taking into account only significant dependencies in between the random variables.

Since the problem of inferring a general network structure from the data is in provably NP-hard, simplifying assumptions are often made about the nature of the existing dependencies between the variables.

In naive Bayesian classifiers, for instance, the considered simplified network consist only of a series of levels of parents and of several mutually independent children nodes. Classification is obtained by considering the parent node to be a hidden variable stating which class each data element should belong to. The classification is termed "naive" since the hypothesis of statistical independence between variables belonging to the same level is not necessarily justified [e.g., see Fig. 1(b)].

When this approximation is considered valid, though, the scene S can be related to the data D through more levels of intermediate simpler models F_i that are causally linked

$$\begin{aligned} &S \rightarrow F_1 \rightarrow \cdots \rightarrow F_n \rightarrow D \quad \text{direct modeling} \\ &S \leftarrow F_1 \leftarrow \cdots \leftarrow F_n \leftarrow D \quad \text{scene understanding.} \end{aligned}$$

The learning is in this case performed across levels, again by making use of the Bayes' equation in a version that considers the various levels of inference involved

$$p(s \mid f_1), \ldots, p(f_n \mid d)p(d) = p(d \mid f_n), \ldots, p(f_1 \mid s)p(s) \quad (3)$$

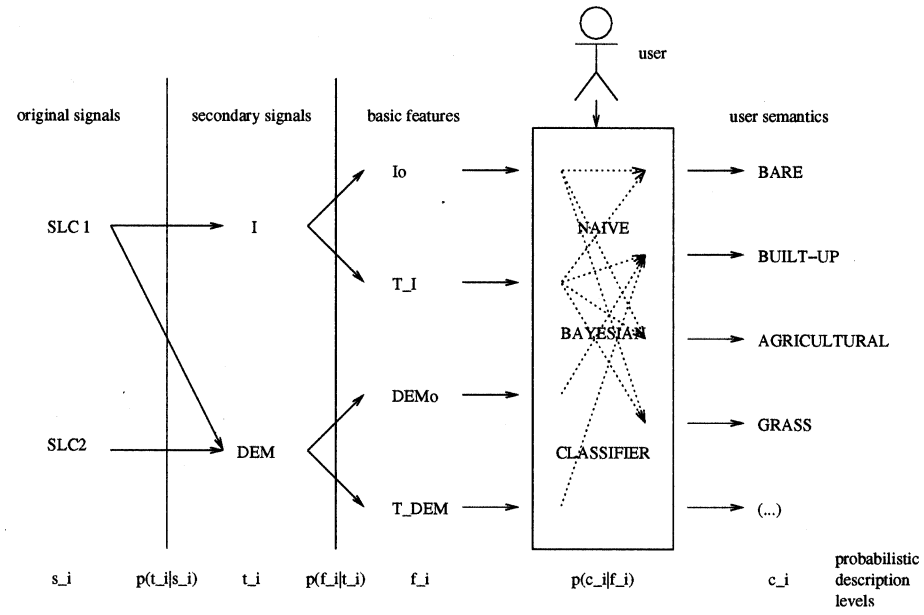
to derive a conclusion about the underlying scene s.

In this framework, a model for the image formation mechanism can be described in the following way:

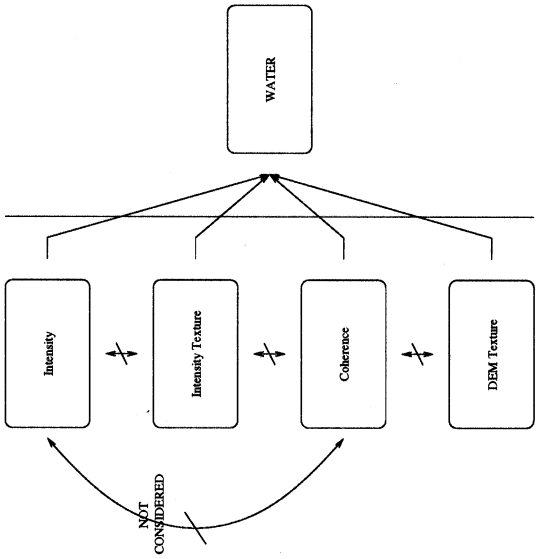
$$\begin{aligned} &S \rightarrow \Omega \rightarrow R \rightarrow D \quad \text{direct model} \\ &S \leftarrow \Omega \leftarrow R \leftarrow D \quad \text{scene understanding.} \end{aligned} \quad (4)$$

A scene S is illuminated and generates a reflected field carrying information in the parameter ω, a realization of the scene feature process Ω. The scene features are translated to signal features R by the instrument and finally to data D by the image formation systems.

The SU procedure must invert this causal chain, recovering the original scene features from the data. The layers of inference needed are implemented in the stages of a processing chain that is described in Fig. 2: the original signals acquired are used to generate image data products (backscatter intensity, coherence and height map) by using standard operational processors. These products are analyzed by a feature extraction system (described in Section III) and translated into a set of image features. The features are grouped by similarity using a simple unsupervised



(a)



(b)

Fig. 1. Learning through hierarchical Bayesian networks in our system. (a) The tree edge structure encodes statistical independence between the random variables in the nodes. The last level of inference is performed by learning the problem semantics through user interaction. As in naive Bayesian classification variables on the same level of the tree structure are assumed mutually independent. This might not be strictly the case, as in the example in (b), which shows that the label WATER might be characterized by very low backscatter intensity and interferometric coherence. Although in general very low backscatter areas also show low coherence values, the relation between the two variables is not taken into account by the classifier.

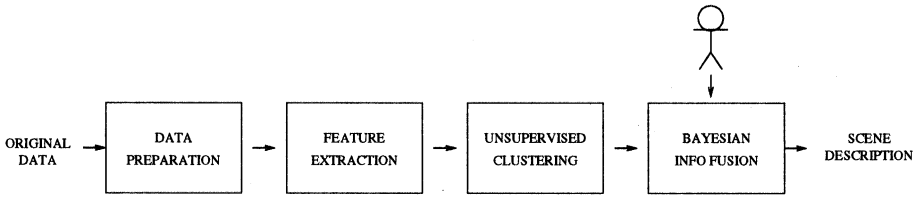


Fig. 2. SU processing chain. Data preparation is followed by information extraction. An unsupervised clustering module performs a grouping of the image features by similarity. Its output is subject to information fusion incorporating semantics by user feedback, thereby producing the final scene description.

classification system, and the generic multidimensional description obtained is fed into a Bayesian classification and information fusion system (Section IV). This new level of description in the data-to-scene chain can incorporate user-provided semantic information into the available representation, finally connecting the generic, multiple-model-based obtained description to the domain of the unknowns in the particular SU problem under investigation.

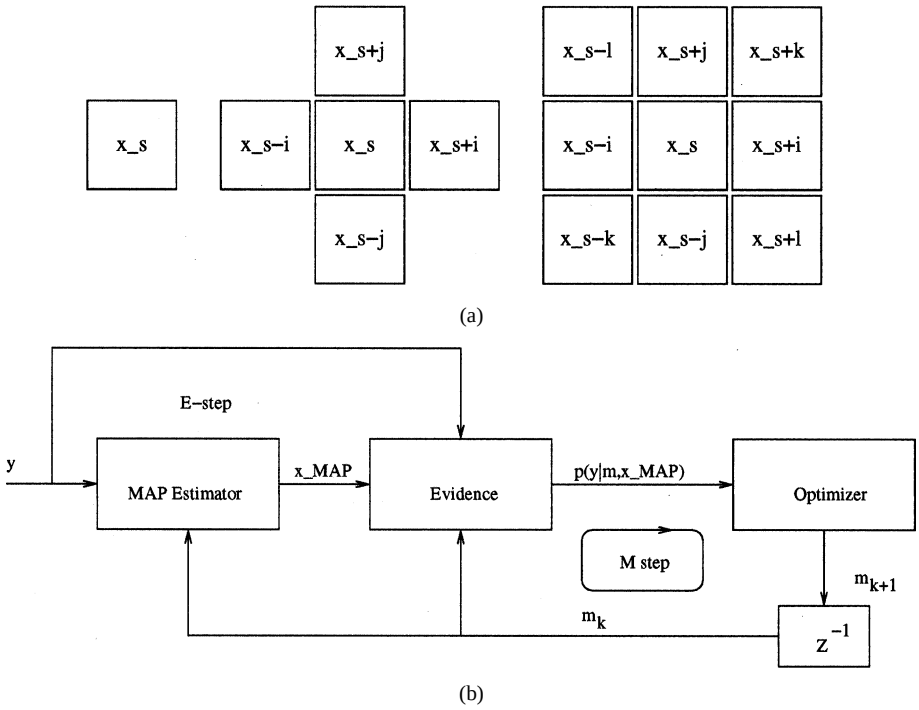


Fig. 3. (a) GMRF clique structures for growing model order (0, 1, 2). (b) Flowchart of the Expectation–Maximization algorithm used for texture parameter estimation and MAP reconstruction in the EMBD despeckling system.

III. Information Extraction From InSAR Data

The feature extraction from SAR intensity and topographic height data is performed respectively by a backscatter despeckling/feature extraction and a DEM filtering/feature extraction subsystem [14], [15].

We postulate in this context that a quasi-complete description of a typical RS image requires a separation of the sources of interesting information and of hindrances in the data—SAR backscatter despeckling, DEM noise filtering—and the estimation of the structural texture properties of the images thereby obtained. The combined feature extraction system models and reconstructs a dataset that is free of noise, while still preserving its most important attributes: it estimates the parameters that describe the data in terms of the employed models. In this sense, it is not only a set of filters, but also a model-based feature extraction system that generates a representation of its input. The model fitting is performed in a space-variant manner that accommodates for local nonstationarities in the data.

Since the system is developed in a Bayesian framework, the choice of appropriate prior models for the clean data plays an important role, and affects in a direct way the obtained results.

The model used as a prior for both backscatter intensity and interferometric DEM analysis is the Gauss–Markov random field (GMRF) [14], [16]

$$p(x_s \,|\, x_{s\pm r}, r \in G, \sigma, m) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_s - \sum_{r\in G} \theta_r x_{s\pm r})^2}{2\sigma^2}\right].$$

It expresses the generic pixel value x_s as normally distributed with width σ around the mean of its neighboring pixels x_{s+r} weighted by the texture parameter vector $\theta = (\theta_0, \ldots \theta_m)$ defined on a neighborhood of cliques G centered on the pixel x_s, and such that the scalar parameters are symmetric around the central element [see Fig. 3(a)].

The main strength of the Gauss–Markov model lies in its ability to model local dependencies between pixels in a wide set of textured images, while still allowing analytical tractability.

While the prior model employed in both information extraction modules is the GMRF, the likelihoods considered have to be adapted to the characteristics of the data under analysis.

The backscatter despeckling and feature extraction subsystem performs a MAP estimation of the radar cross section by considering a space-variant square root Gamma-shaped likelihood

$$p_{\sigma_0}(y_i \,|\, x_i) = 2\left(\frac{y_i}{x_i}\right)^{2L-1} \frac{L^L}{x_i \Gamma(L)} \exp\left[-L\left(\frac{y_i}{x_i}\right)^2\right]$$

where y_i and x_i are the corrupted and the original backscatter square-root intensity values; L is the number of looks of the data; and $\Gamma(\cdot)$ is the Gamma function.

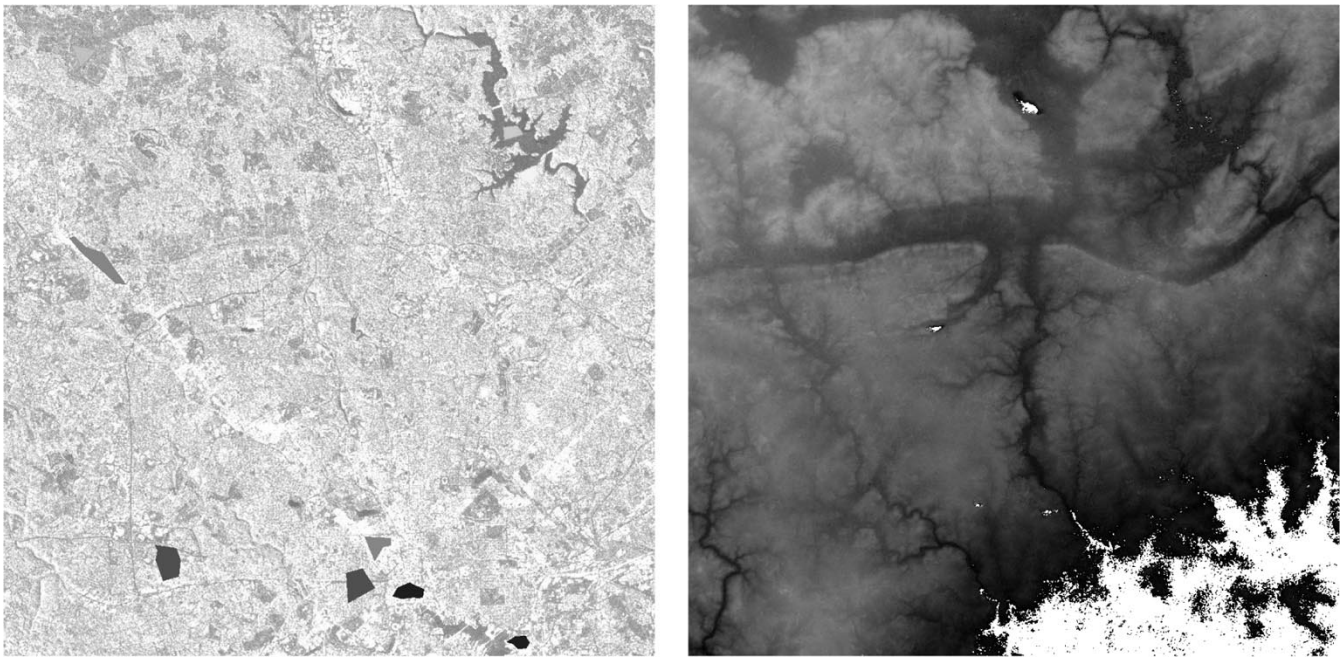
These estimates are then employed to produce parameters for the texture random field by Expectation–Maximization [19] [see Fig. 3(b)] of the analytical evidence term $p(y_i)$.

Nonlinear features such as sharp edges and targets are also extracted from the data and separately handled in order to restore them in the filtered image.

The DEM information extraction system needs to consider the peculiarities of additive Gaussian noise in its likelihood

$$p_{\hat{h}}(y_i \,|\, x_i) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{(y_i - x_i)^2}{2\sigma_n^2}\right]$$

where σ_n is the width of the distribution, again estimated via Expectation–Maximization. It outputs a clean DEM and a local DEM texture description.



(a) (b)

Fig. 4. (a) SRTM backscatter intensity image (resolution 25 m, number of looks four) on the city of Baltimore, MD. A strong multiplicative noise called speckle generated from the coherent nature of the SAR system is visible. Training for supervised classification/feature fusion is superimposed on the data: with growing luminosity from dark to light gray (upper right) AGRICULTURAL, (upper right) WATER, (upper left) FOREST, (center low) MIDDLE_URBANIZED, (left low and center) SPARSE_URBANIZED, (right low and center) DENSE_URBANIZED. (b) corresponding SRTM height image on the city of Baltimore. Artifacts on unmasked water areas are visible.

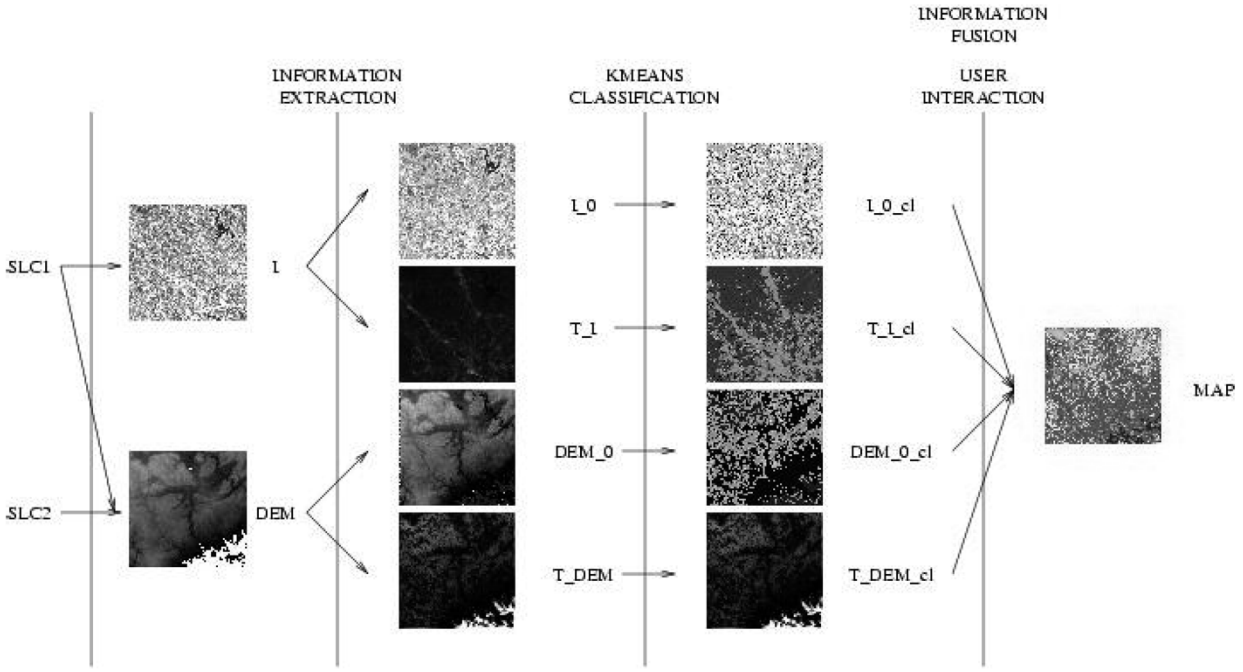


Fig. 5. SU procedure for urban land use mapping on the Baltimore scene with intermediate and final processing results. The original backscatter intensity and elevation data are processed into their clean estimates and textural descriptions through the information extraction system. The features obtained are grouped by similarity and fed into the Bayesian classification/fusion system together with the training in Fig. 4(a) to obtain the final land use map.

IV. Information Fusion and Machine Learning for RS Scene Understanding

After the basic features have been extracted by the combined image information extraction system, they are grouped by similarity, by using a simple unsupervised K-means classification system. The multiple layers of information thereby obtained have then to be fused with each other: the clusters—here indicated as ω_i—play the role of an abstract image vocabulary that is able to explain, by different combinations, the image semantics S.

Instead of imposing an explicit definition of the phenomenon under study as in a rule-based expert system, descriptions are learned from the human user by example. This implies a much

more direct and powerful way of providing information to the system and enabling it to consider the interpreter conjectures.

Again, to be able to capture the subjective interpreter-dependent aspects of information, a Bayesian formalization is needed: probability is interpreted as a degree of belief rather than as a frequency of realization. This contextual fusion and interactive classification is performed via a Bayesian classification and fusion system.

The process of interactive learning consists in the possibility of progressively computing the probability of a specific cover-type in the image, on the basis of positive and negative pixel-level examples provided by the user. The inference process for the label $S = s_\nu$ given the image data d and the features in the clusters ω_i is realized through the probability

$$p(s_\nu \,|\, d) = p(s_\nu) \sum_i \frac{p(\omega_i \,|\, s_\nu) p(\omega_i \,|\, d)}{p(\omega_i)}$$

where $p(s_\nu)$ and $p(\omega_i)$ are (usually noninformative) prior probabilities, while $p(\omega_i \,|\, s_\nu)$ has to be learnt from user examples. An independence condition is assumed, i.e., the probability is assumed equal to the product of the separate likelihoods for the cover type d_ν given each single model

$$\begin{aligned} p(\omega_i \,|\, s_\nu) &= p\left(\omega_{i,M_1}, \omega_{i,M_2}, \ldots, \omega_{i,M_m} \,|\, s_\nu\right) \\ &= p\left(\omega_{i,M_1} \,|\, s_\nu\right) p\left(\omega_{i,M_2} \,|\, s_\nu\right) \ldots p\left(\omega_{i,M_m} \,|\, s_\nu\right) \end{aligned}$$

since each cover type is a considered a combination of different models

$$\omega_i = \omega_{M_1} \otimes \omega_{M_2} \otimes \cdots \otimes \omega_{M_m}.$$

If we denote with ϕ the vector parameter that satisfies the identities

$$p(\omega_i \,|\, s_\nu, \phi) = \phi_i, \quad p(\phi) = \Gamma(r)$$

where $\Gamma(\cdot)$ is the Gamma distribution, then the probability of the training set is

$$p(\phi \,|\, T) = \mathrm{Dir}(\phi \,|\, 1 + N_1, \ldots, 1 + N_r)$$

where $\mathrm{Dir}(\cdot)$ is the Dirichlet distribution, and N_i is the number of occurrences of the signal type ω_i in T. If a new training set T' is provided, the probability is updated according to

$$p(\phi \,|\, T, T') = \mathrm{Dir}(\phi \,|\, 1 + N_1 + N_1', \ldots, 1 + N_r + N_r').$$

Denoting with a the parameter vector that satisfies the identity

$$a_i = 1 + N_i$$

the learning process is modeled by updating the vector a after observation of each training set.

V. Urban Land Use Mapping From SRTM Data

We apply the whole procedure described—and summarized in Fig. 2—to urban land use mapping on the city of Baltimore, MD, from an SRTM dataset. A U.S. Geological Survey (USGS) map of the scene is reported in Fig. 6(b), while the original backscatter and the DEM data are shown in Fig. 4. No interferometric coherence information is used.

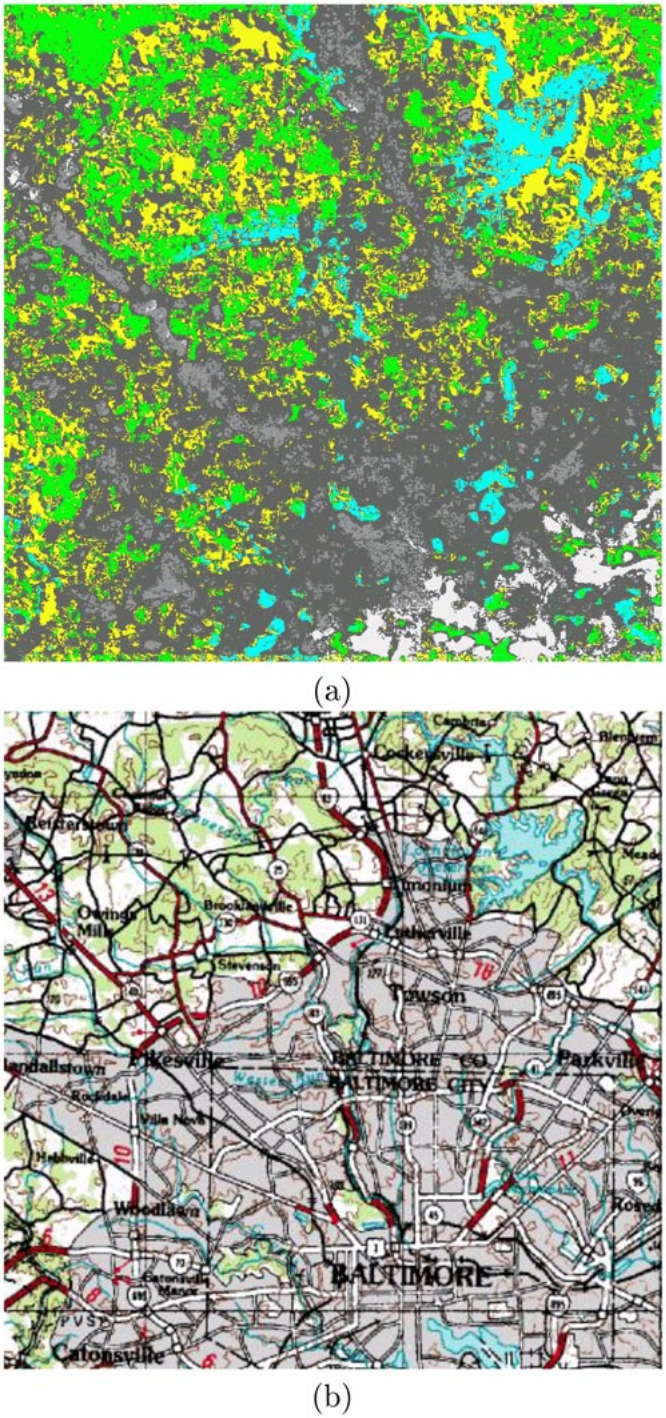


Fig. 6. (a) Bayesian classification/fusion results. FOREST (green), WATER (light blue), SPARSE_URBANIZED (middle gray), MIDDLE_URBANIZED (light gray), DENSE_URBANIZED (very light gray), AGRICULTURAL (yellow). (b) USGS July 1, 1984 ground truth (from http://terraserver.homeadvisor.msn.com/). The main directions of development of the city are visible as radial areas of higher building density, while the city center and the areas with highest urbanization around it are separated from the rest of the urbanized regions. Misclassifications appear on areas where the DEM information is corrupted by artifacts.

Before being fed into the Bayesian classification and information fusion system, the original dataset layers are subject to separate information extraction and unsupervised classification. A description of the procedure and the intermediate results are shown in Fig. 5.

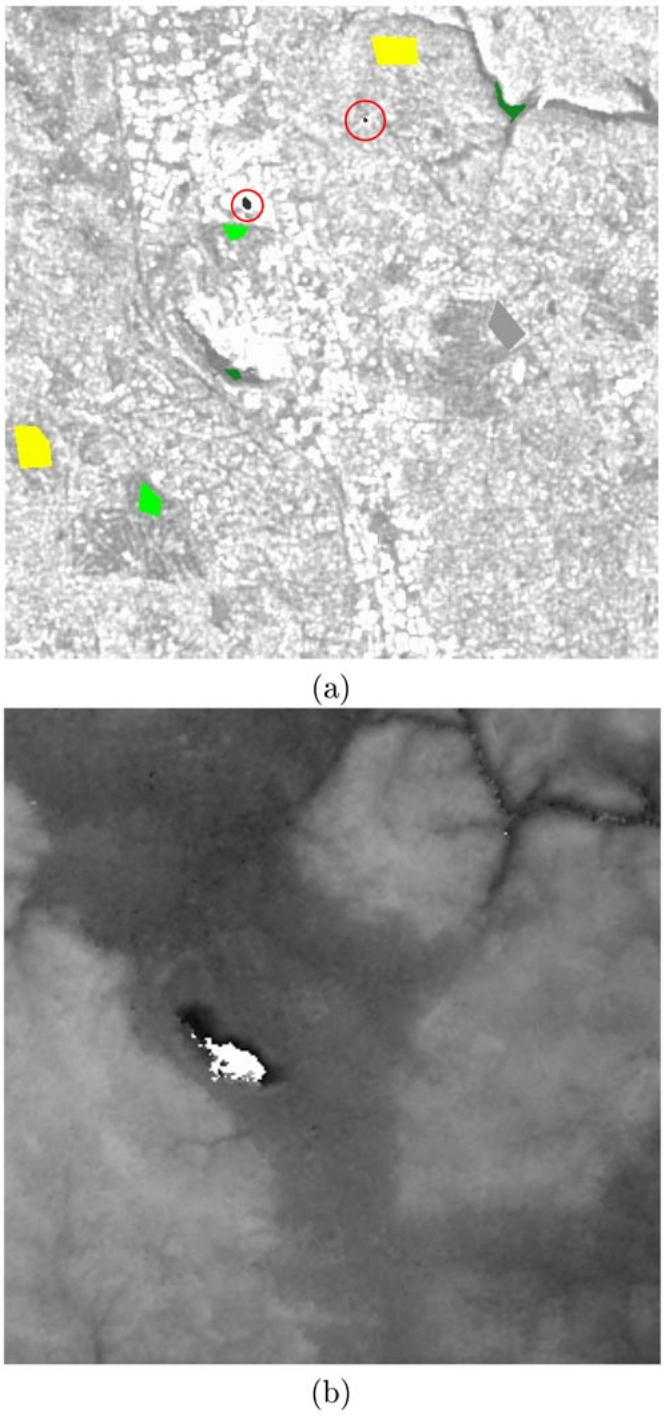


Fig. 7. (a) SRTM backscatter intensity image on the area object of building recognition with superimposed Bayesian classification/fusion training. BARE-AGRICULTURAL (yellow), BUILDING (dark gray in red circles), WATER/SHADOW (dark green), GENERAL_URBANIZED (light gray), FOREST (light green). (b) SRTM height image on the same area. Artifacts are visible on unmasked shadow/water areas.

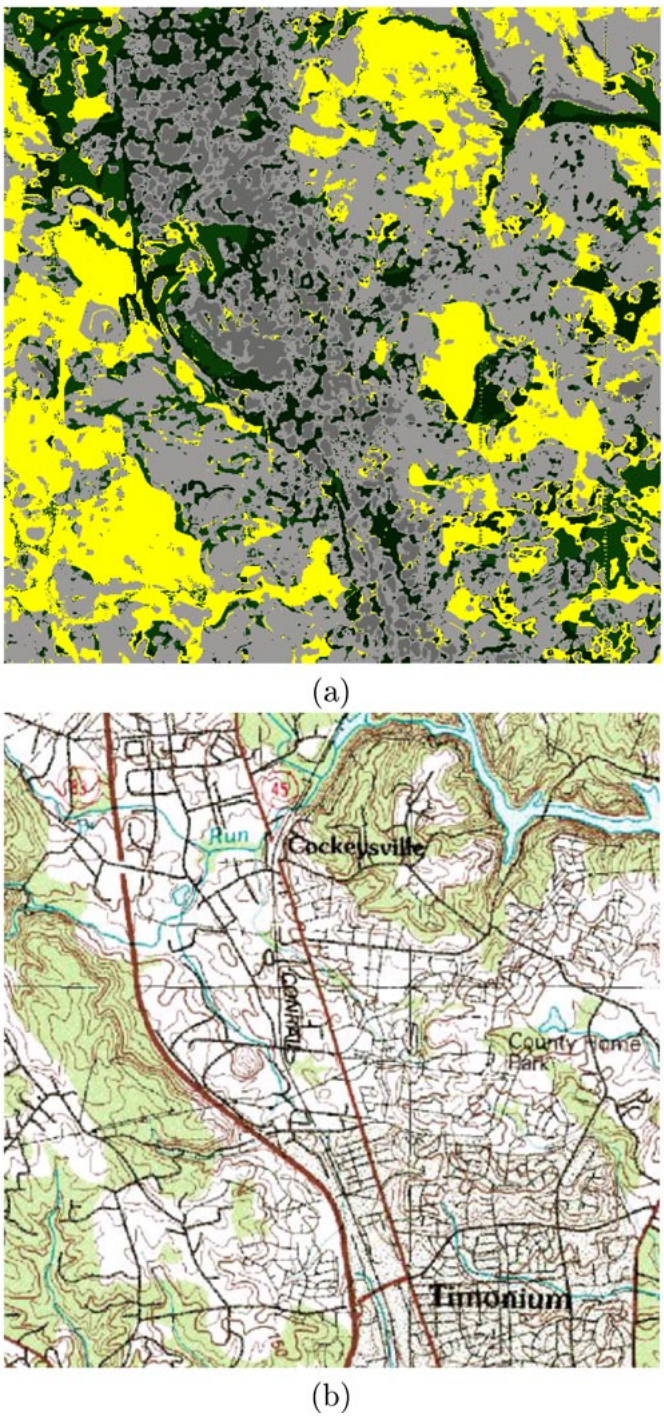


Fig. 8. (a) Bayesian classification/fusion results. BARE-AGRICULTURAL (yellow), BUILDING (dark gray), WATER/SHADOW (dark green), GENERAL_URBANIZED (light gray), FOREST (light green). (b) USGS July 1, 1984 ground truth (from http://terraserver.homeadvisor.msn.com/).

The information fusion step takes user input in the form of the definition of training regions. The ones defined for the scene under study are reported in Fig. 4(a) together with their class definitions. The obtained classified results are reported in Fig. 6(a) together with USGS ground truth: the main directions of development of the city are visible as radial areas of higher building density, while the city center and the areas with highest urbanization around it are well separated from the rest of the urbanized regions. Limited misclassifications are visible in the lower right part of the dataset where the original height information is corrupted by strong artifacts generated by unmasked water areas.

VI. Building Recognition From SRTM Data

Although the spatial resolution (25 m) of the SRTM dataset limits its applicability to the understanding of very complex

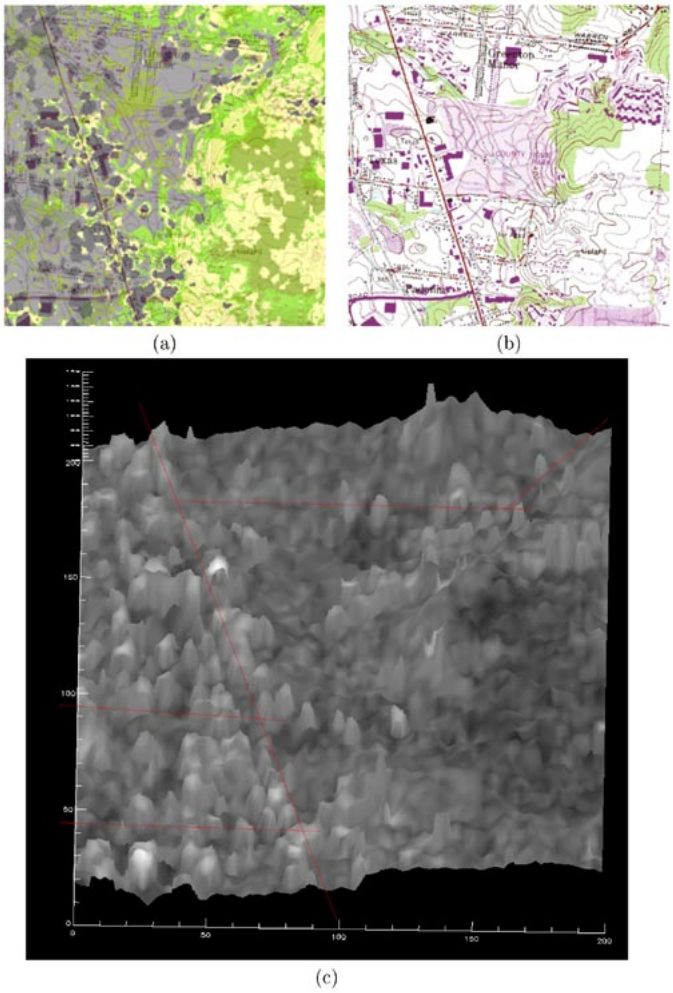


Fig. 9. (a) Bayesian classification/fusion results with ground truth in transparency. (b) USGS July 1, 1986 ground truth (from http://terraserver.homeadvisor.msn.com/). All but one of the buildings reported on the 1986 map are recognized as such. The remaining detected buildings are either actual buildings erected after the ground truth map was produced or false alarms. (c) Three-dimensional view of center portion (about 2.5 km of side) of the local DEM where the heights of class BUILDING have been exaggerated by 10% with superimposed annotation data. The distribution of some of the streets in the scene (superimposed in red) is visible.

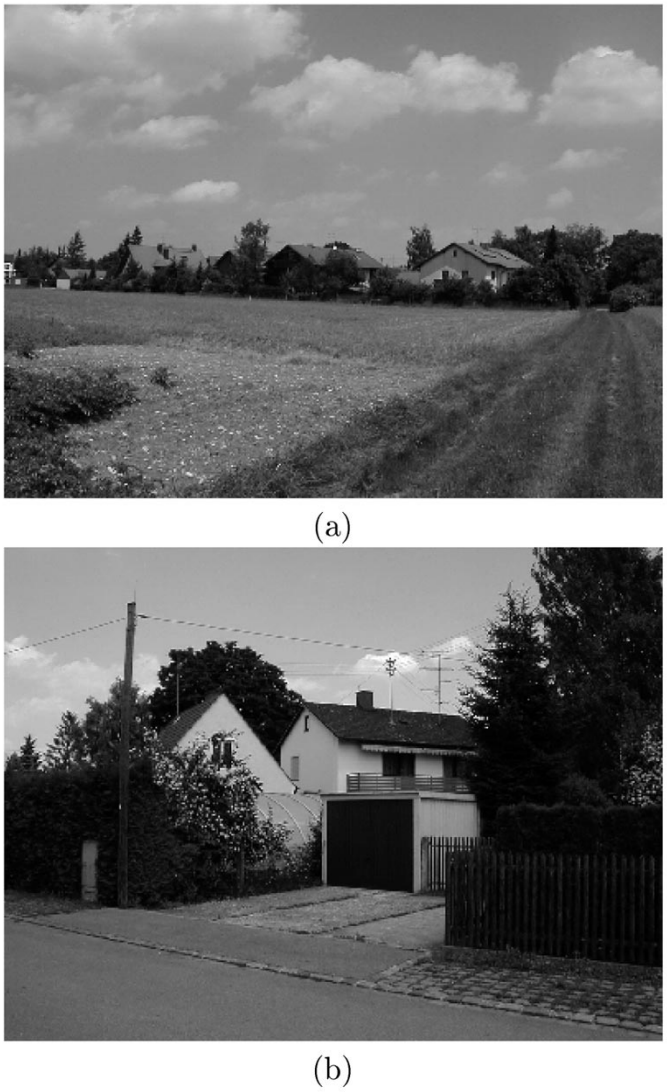


Fig. 10. Views of the rural test area. Buildings are partially hidden by trees at X-band. Viewpoints are indicated in Fig. 12.

urban environments, we evaluate the results of the application of the described SU framework to the identification of large buildings.

The area of the city that we concentrate our analysis on (about 10×10 km wide) is reported in Fig. 8(b): it includes a city park as well as highly developed urban elements. Original backscatter intensity and heights are reported in Fig. 7.

The input training regions provided to the interactive Bayesian information fusion system are described in Fig. 7(a). The results of the procedure are reported in Fig. 8: the different elements in the scene are discriminated from each other, and many individual buildings are separated.

A perspective view of center portion (about 2.5 km of side) of the local DEM where the heights of class BUILDING have been exaggerated by 10% [in Fig. 9(c)] shows the distribution of some of the streets in the scene.

A comparison with the ground truth in Fig. 9(b) shows that all but one of the buildings reported in the 1986 USGS ground truth map are recognized as such. The remaining detected buildings, many of which are aligned to the existing road network, are either actual buildings erected after the ground truth map was produced or false alarms.

VII. Building Recognition From 0.5-m Resolution Intermap Data

While the resolution of the SRTM system makes it more appropriate for the study of large-scale natural structures than for the understanding of complex urban environments, the next generation of meter-resolution spaceborne SAR sensors will be producing on a regular basis vast amounts of detailed data on urban environments.

Despite differences in some aspects of the sensor characteristics, current airborne systems provide a test bed for the evaluation of SU algorithms to next-generation spaceborne data.

An Intermap X-band interferometric dataset with resolution of about 0.5 m taken during a flight on a rural area east of the city

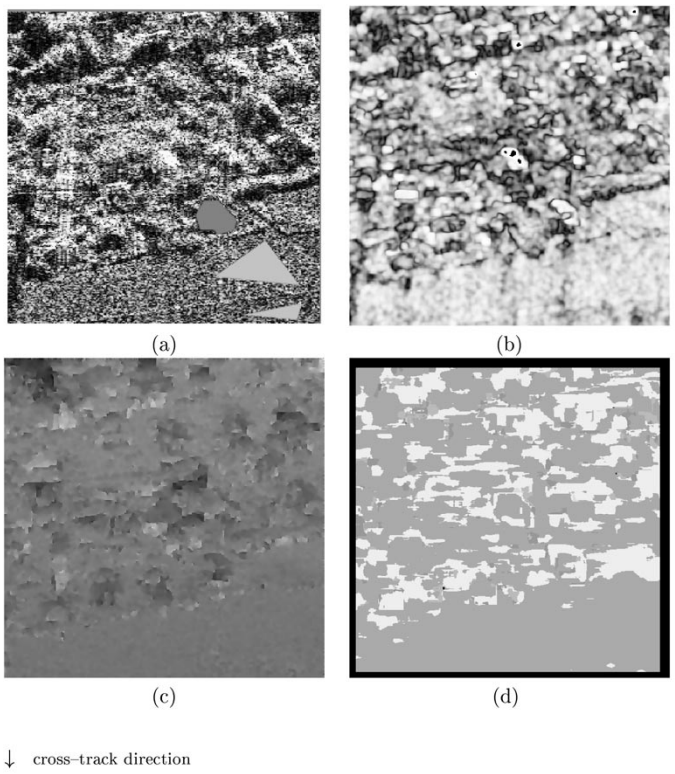


Fig. 11. Rural test scene data. (a) Intensity image of the site presented in Fig. 10. The very limited training areas are chosen to allow the detection of buildings by separating them from other scene structures. The information on buildings is extracted from intensity image, (b) coherence, (c) interferometry-derived elevations, and (d) interferometric phase gradients, and further fused according to the "example" given by the training area. The cross-track direction is indicated as an aid in the interpretation of the data, given the strong dependence of meter-resolution SAR phenomenology on the acquisition geometry.

of Munich, Germany, is used in a further building recognition experiment.

The test scene has very high complexity. The structures to be reconstructed are small houses partially hidden at X-band by surrounding taller vegetation (see Fig. 10). Typically for very high-resolution SAR data in built-up environments, multiple scattering and other disturbing effects tend to dominate the data, saturating the receiver and propagating to different areas in the image (Fig. 11).

The signal features used include the interferometric coherence, interferometric phase gradients obtained via Gabor filters and single image intensities and intensity textures as well as interferometry-derived measures of the local elevations. The signal features are separately subject to feature extraction and unsupervised classification and are subsequently fused in a supervised way after specifying the training mask in Fig. 11(a).

Although the user-provided training is intentionally limited to two very simple areas describing respectively built-up and natural scene elements, the results tend to separate built-up elements from different image objects such as trees or vegetation. Buildings in the scene are marked as such, while some spurious classification in the middle of the scene is determined by responses by greenhouses and other nonpermanent structures. The results obtained are shown together with ground truth in Fig. 12.

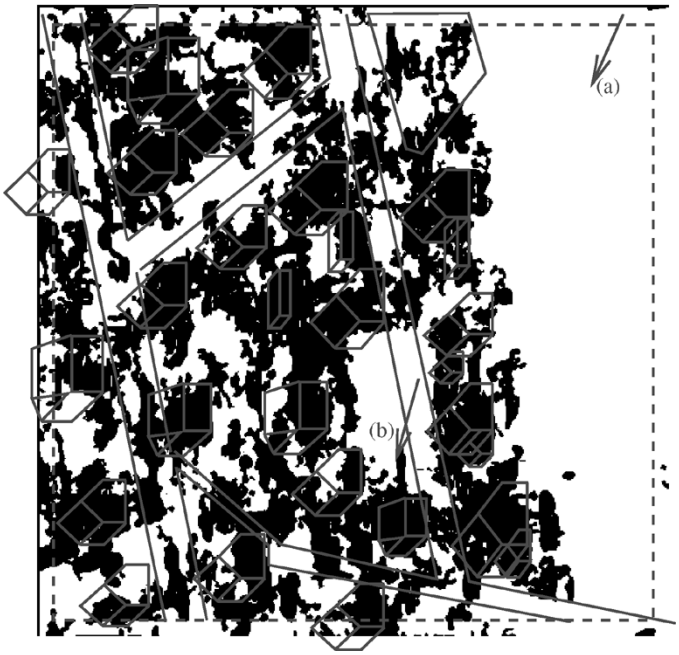


Fig. 12. Rural test scene results. (a) Extracted building map with ground truth provided for comparison. The arrows mark the view points for the photographs in Fig. 10, while the box marks invalid results on image borders.

VIII. Conclusion

We have presented a framework for scene understanding from InSAR data that is based on Bayesian machine learning and information extraction and fusion. The procedure is applicable at multiple scales. We have given examples on data with spatial resolutions in the range 25–0.5 m obtained from both airborne and spaceborne sensors with applications ranging from urban land use map generation to building recognition. The same procedures have also been applied to further datasets acquired through a number of different sensors including X-SAR and E-SAR.

Acknowledgment

The high-resolution rural test scene data were provided by Intermap Technologies GmbH. The authors would also like to thank M. Ciucu, A. Fusco, and A. Pelizzari for their useful comments.

References

[1] A. Huertas, Z. Kim, and R. Nevatia, "Use of IFSAR with intensity images for automatic building modeling," DARPA, Los Angeles, CA, DARPA Project Annual Rep., 1998.
[2] R. Bolter and F. Leberl, "Phenomenology-based and interferometry-guided building reconstruction from multiple SAR images," in *Proc. EUSAR 2000*, Munich, Germany, 2000, pp. 687–690.
[3] C. Prati, F. Rocca, A. M. Guarnieri, and P. Pasquali, "Report on ERS-1 SAR Interferometric Techniques and Applications," ESA, Tech. Rep., ESA Study Contract Rep., 1994.
[4] N.A.C. Cressie, *Statistics for Spatial Data*. New York: Wiley, 1991.
[5] J. C. Curlander and R. N. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing*. New York: Wiley, 1992.
[6] D. Hackerman, D. Geiger, and D. M. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data," *Mach. Learning*, vol. 20, no. 3, pp. 197–243, 1995.

[7] E. Simonetto, H. Oriot, and R. Garello, "Extraction of industrial structures and DEM from airborne SAR images," in *Proc. 2nd Int. Symp., PSIP'2001*, Marseille, France, Jan. 23–24, 2001, pp. 412–417.
[8] F. dell'Acqua and P. Gamba, "Preparing an urban test site for SRTM data validation," *IEEE Trans. Geoscience Remote Sensing*, vol. 40, pp. 2248–2256, Oct. 2002.
[9] F. Tupin, H. Maitre, J.-F. Mangin, J. M. Nicolas, and E. Pechersky, "Detection of linear features in SAR images: Application to road network extraction," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 434–453, Mar. 1998.
[10] P. Gamba, F. dell'Acqua, and B. Houshmand, "SRTM data characterization in urban areas," in *Proc. Photogrammetric Computer Vision ISPRS Commission III, Symp. 2002*, Graz, Austria, Sept. 9–13, 2002.
[11] P. Gamba and B. Houshmand, "Three-dimensional road network by fusion of polarimetric and interferometric SAR data," in *Proc. IGARSS*, vol. I, Hamburg, Germany, June 1999, pp. 302–304.
[12] G. Franceschetti, A. Iodice, and D. Riccio, "A canonical problem in electromagnetic backscattering from buildings," *IEEE Trans. Geosci. Remote Sensing*, vol. 40, pp. 1787–1801, Aug. 2002.
[13] G. R. Burkhart, Z. Begen, and R. Carande, "Elevation correction and building extraction from interferometric SAR imagery," in *Proc. IGARSS*, 1996, pp. 659–660.
[14] M. Datcu, K. Seidel, and M. Walessa, "Spatial information retrieval from remote sensing images—Part I: Information theoretical perspective," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 1431–1445, Sept. 1998.
[15] M. Datcu and M. Walessa, "Model-based despeckling and information extraction from SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 2258–2269, Sept. 2000.
[16] M. Schröeder, H. Rehrauer, K. Seidel, and M. Datcu, "Spatial information retrieval from remote sensing images—Part II: Gibbs Markov fields," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 1446–1455, Sept. 1998.
[17] P. Gamba and B. Houshmand, "Three dimensional urban characterization by IFSAR measurements," in *Proc. IGARSS*, 1999, pp. 2401–2403.
[18] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.
[19] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Mag.*, vol. 11, pp. 47–60, 1996.
[20] R. Touzi, A. Lopes, J. Bruniquel, and P. W. Vachon, "Coherence estimation for SAR imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 135–149, Jan. 1999.

Marco Quartulli received the laurea degree in physics from the University of Bari, Bari, Italy, in 1997 with a thesis on differential SAR interferometry. He is currently purusing the Ph.D. degree.

From 1997 to November 2000, he worked on SAR and InSAR processing systems at Advanced Computer Systems, Rome, Italy. In 2000, he joined the Image Analysis Group at the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germnay. His research interests include Bayesian modeling and scene understanding for high-resolution SAR and InSAR.

Mihai Datcu received the M.S. and Ph.D. degrees in electronics and telecommunications from the University "Politehnica" of Bucharest UPB, Bucharest, Romania, in 1978 and 1986, and the title "Habilitation a diriger des recherches" from Université Louis Pasteur, Strasbourg, France.

He holds a Professorship in electronics and telecommunications with UPB since 1981. Since 1993, he has been a Scientist with the German Aerospace Center (DLR), Oberpfaffenhofen, Germany. He is currently developing algorithms for model-based information retrieval from high-complexity signals, methods for scene understanding from SAR and interferometric SAR data, and he is engaged in research in information theoretical aspects and semantic representations in advanced communication systems. He has held Visiting Professor appointments from 1991 to 1992 with the Department of Mathematics, University of Oviedo, Oviedo, Spain, from 2000 to 2002 with the Université Louis Pasteur, and the International Space University, Strasbourg, France. In 1994, he was a Guest Scientist with the Swiss Center for Scientific Computing (CSCS). Manno, Switzerland, and in 2003, he was a Visiting Professor with the University of Siegen, Siegen, Germany. From 1992 to 2002, he had longer Visiting Professor assignments with the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland. He is involved in advanced research programs for information extraction, data mining and knowledge discovery, and data understanding with the European Space Agency, Centre National d'Etudes Spatiales, the National Aeronautics and Space Administration, and in a variety of European projects. He is currently Senior Scientist and Image Analysis research group leader with the Remote Sensing Technology Institute (IMF), DLR. His research interests are in Bayesian inference, information and complexity theory, stochastic processes, model-based scene understanding, image information mining, for applications in information retrieval and understanding of high-resolution SAR and optical observations.