Discovery of Semantic Relationships in PolSAR Images Using Latent Dirichlet Allocation

Radu Tănase, Reza Bahmanyar, Gottfried Schwarz, and Mihai Datcu, *Fellow, IEEE*

***Abstract*— We propose a multilevel semantics discovery approach for bridging the semantic gap when mining high-resolution polarimetric synthetic aperture radar (PolSAR) remote sensing images. First, an Entropy/Anisotropy/Alpha-Wishart classifier is employed to discover low-level semantics as classes representing the physical scattering properties of targets (e.g., low-entropy/surface scattering/high anisotropy). Then, the images are tiled into patches and each patch is modeled as a bag-of-words, a histogram of the class labels. Next, latent Dirichlet allocation is applied to discover their higher level semantics as a set of topics. Our results demonstrate that topic semantics are close to human semantics used for basic land-cover types (e.g., grassland). Therefore, using the topic description (bag-of-topics) of PolSAR images leads to a narrower semantic gap in image mining. In addition, a visual exploration of the topic descriptions helps to find semantic relationships, which can be used for defining new semantic categories (e.g., mixed land-cover types) and designing rule-based categorization schemes.**

***Index Terms*— Bag-of-topics (BoT), bag-of-words (BoW), Entropy/Anisotropy/Alpha-Wishart classification, latent Dirichlet allocation (LDA), polarimetric synthetic aperture radar (PolSAR), semantic relationships.**

I. Introduction

THE multitude of modern remote sensing sensors allows us to analyze tremendous amounts of high-resolution earth observation (EO) images. Therefore, developing new content-based image retrieval (CBIR) systems being able to extract user-desired information from existing image databases is highly demanded. In the state-of-the-art literature, various CBIR systems have been proposed for EO image mining, such as Intelligent Interactive Image Knowledge Retrieval System (I^3KR) [1], Knowledge-Driven Information Mining System [2], and its accelerated variant [3]. For the existing systems, semantic image interpretations are usually provided through either manual image annotation or user acceptance (in active learning scenarios), which require much human effort and time, and bias the systems toward user perspectives [4]. Moreover, due to the differences between human image understanding and how computers interpret and process them (the so-called *semantic gap* [4], [5]), many image mining results provided by computers are still unsatisfactory.

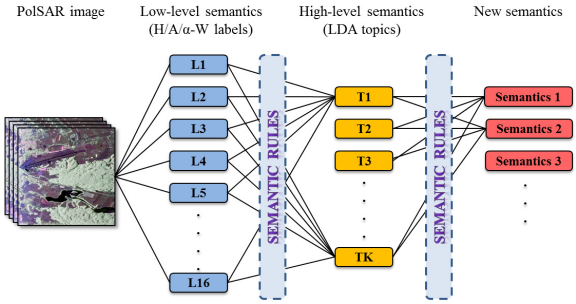

Fig. 1. Stepwise low- to high-level semantics generation.

In this letter, we propose a multilevel semantics discovery approach for bringing computer interpretation of a particular remote sensing image type, namely, *polarimetric synthetic aperture radar* (PolSAR) images, closer to human semantics. This helps computers to discover existing semantic relationships within images and employ them for analyzing user queries (which are based on semantics) and provide them with semantically meaningful relevant results. Fig. 1 shows an overview of the proposed approach.

PolSAR images supply information with respect to the physical scattering properties of the recorded ground targets, retrieved by applying coherent or incoherent target decomposition theorems to the first- and second-order polarimetric representations [6]. A widely employed decomposition method is *Entropy/Anisotropy/Alpha* (H/A/α), resulting in three parameters describing the physics behind the scattering processes. These parameters lead to a superior pixel-based unsupervised classification scheme, the H/A/α classification method [7]. A modification to this method, namely, the *H/A/α-Wishart* classification method [8], shows that the complex Wishart distribution parameters improve the classification substantially.

In our proposed approach, we employ the H/A/α-Wishart method for discovering the low-level semantics of PolSAR images as a set of classes, representing targets by their physical scattering properties (e.g., low-entropy surface scattering (SS) with high anisotropy). The images are then tiled into patches, and each image patch is modeled as a *bag-of-words* (BoW) by generating a histogram of its assigned class labels, where the labels are the words in the BoWs (see Fig. 2). After that, a generative statistical model, namely, *latent Dirichlet allocation* (LDA) [9], is applied to the BoW histograms in order to discover the latent semantics behind the image patches

Manuscript received July 26, 2016; revised October 5, 2016 and November 28, 2016; accepted November 30, 2016. Date of publication December 23, 2016; date of current version January 19, 2017.

R. Tănase is with the Research Centre for Spatial Information, CEOSpaceTech, University Politehnica of Bucharest, 060042 Bucharest, Romania, and also with the Military Technical Academy, 050141 Bucharest, Romania (e-mail: radu.tanase@ceospacetech.pub.ro).

R. Bahmanyar and G. Schwarz are with the German Aerospace Center, Remote Sensing Technology Institute, Wessling 82234, Germany (e-mail: reza.bahmanyar@dlr.de; gottfried.schwarz@dlr.de; mihai.datcu@dlr.de).

M. Datcu is with the German Aerospace Center, Remote Sensing Technology Institute, Wessling 82234, Germany, and also with the Research Centre for Spatial Information, CEOSpaceTech, University Politehnica of Bucharest, 011061 Bucharest, Romania (e-mail: mihai.datcu@dlr.de).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2016.2636663

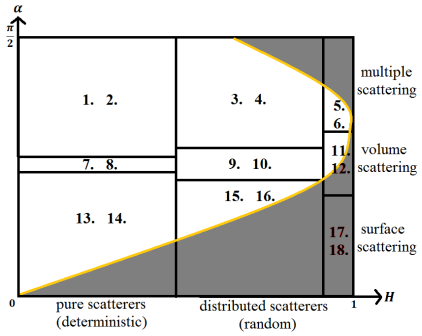

Fig. 2. H/α classification scheme (adapted from [7]). From the 18 classes, only 1–16 are feasible and are assigned as labels $L1$–$L16$ in Table I.

as a set of topics. Our validation of the topics based on ground truth (Google Earth[1]) images demonstrates that they provide high-level semantics, which are close to human semantics used for identifying land-cover types (e.g., woody vegetation).

As the topics are the basic land-cover types existing in the image patches, further land-cover types can be defined as combinations of the basic land-cover types (e.g., a shoreline is a combination of a water body, grassland, and woody vegetation). Thus, the image patches can be modeled by vectors of topic mixtures, the co-called *bag-of-topics* (BoT) model [10]. The topic vectors form a multidimensional Euclidean space in which each image patch is represented as a point. Since the dimensions of this space are semantically meaningful, it can be easily explored and assessed through immersive visualization techniques to discover existing semantic relationships and identify new semantic categories (e.g., mixed land-cover types). While a topic representation of images can adapt computer image interpretation to human semantics, the semantic relationships can be used for designing rule-based land-cover categorization methods.

Section II reviews H/A/α-Wishart classification. Section III describes low-level semantics discovery. Section IV briefly introduces LDA. Sections V and VI explain how to discover high-level semantics and explore it for finding new semantics. Section VII concludes this letter.

II. Entropy/Anisotropy/Alpha-Wishart Classification

Our test image is an F-Synthetic Aperture Radar System airborne complex-valued data set[2] comprising four polarization planes (VV, HH, HV, and VH) [11]. The data were multilooked with a factor of 5 in azimuth direction. In order to reduce the inherent speckle noise, PolSAR data are usually delivered in a multilooked pixel-wise coherency matrix format T. This matrix is a 3×3 Hermitian, positive, and semidefinite matrix, which can be written as

$$T = U \cdot \Sigma \cdot U^{-1} \tag{1}$$

where Σ is a diagonal matrix composed of the eigenvalues of T (λ_1, λ_2, and λ_3 in descending order). The columns of U contain the corresponding eigenvectors to the eigenvalues (u_1, u_2, and u_3), where each u_i can be further decomposed into [6]

$$u_i = \begin{bmatrix} \cos\alpha_i & \sin\alpha_i \cos\beta_i e^{j\delta_i} & \sin\alpha_i \cos\beta_i e^{j\gamma_i} \end{bmatrix}^T. \tag{2}$$

[1]https://www.google.com/earth/

[2]courtesy of Rolf Scheiber from DLR's Microwaves and Radar Institute

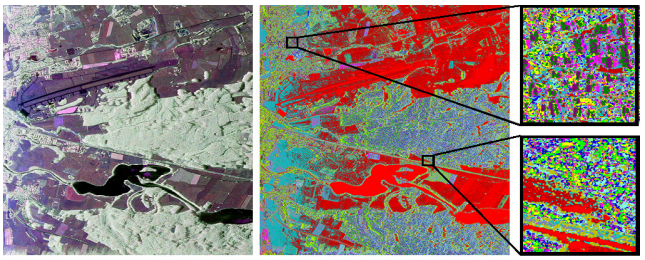

Fig. 3. False-color Pauli RGB representation of our test image (left). H/A/α-Wishart classification map together with two zoomed-in subareas (right), where the colors correspond to the low-level semantics in Table I.

From (1) and (2), the following parameters can be computed:

$$H = \sum_{i=1}^{3} -p_i \log_3 p_i, \quad A = \frac{p_2 - p_3}{p_2 + p_3}, \quad \alpha = \sum_{i=1}^{3} p_i \alpha_i \tag{3}$$

where $p_i = \lambda_i / \sum_{j=1}^{3} \lambda_j$ denotes the probability of the eigenvalues λ_i. These parameters refer to the physics behind the scattering processes. The entropy H discriminates pure and distributed scatterers; the anisotropy A characterizes different types of scattering [6], and the mean α angle shows the dominant scattering mechanism.

Combinations of these parameters can lead to very good classification schemes. For example, the scheme in Fig. 2 divides the H/α plane into nine zones (classes), from which only eight are feasible. Including anisotropy will double the number of classes: nine for $A \leq 0.5$ (denoted by odd labels) and nine for $A > 0.5$ (denoted by even labels). The $H/A/\alpha$ classification method has its own drawbacks, such as the fixed boundaries of the classes, which do not dynamically adapt to the input data. Dealing with this issue, Lee *et al.* [8] proposed to use the parameters of the complex Wishart distribution of the coherency matrix. This approach employs the $H/A/\alpha$ classification for initializing the classes, and then uses the Wishart parameters in an iterative procedure to refine the classification. Its main advantages are to consider the magnitude of the coherency matrix, which is important in detecting SS, and allowing a dynamic adaptation of the class boundaries to the input data.

III. Low-Level Semantics Discovery

We apply the $H/A/\alpha$ – Wishart classification, based on 5×5 pixel windows, to our PolSAR test image, shown in Fig. 3 (left), in order to retrieve a classification map with 16 classes. The class labels are then considered as low-level semantics, which refer to the scattering properties of the recorded targets. Fig. 3 (right) shows the classification results, with Table I defining the color coding.

The labels in Table I can be categorized into three main scattering mechanisms: the first six labels ($L1$–$L6$) refer to *multiple scattering* (MS), which usually occur in urban and forested areas. The next six labels ($L7$–$L12$) refer to *volume scattering* (VS), which is usually observed in forested and vegetated areas, while the labels $L13$–$L16$ refer to SS, which can be usually seen on rough land surfaces. These scattering mechanisms can be further divided into several subcategories depending on H, which hints to the number of scatterers

TABLE I
LOW-LEVEL SEMANTICS

Label	Semantics
L1	Low-entropy multiple scattering with low anisotropy
L2	Low-entropy multiple scattering with high anisotropy
L3	Medium-entropy multiple scattering with low anisotropy
L4	Medium-entropy multiple scattering with high anisotropy
L5	High-entropy multiple scattering with low anisotropy
L6	High-entropy multiple scattering with high anisotropy
L7	Low-entropy volume scattering with low anisotropy
L8	Low-entropy volume scattering with high anisotropy
L9	Medium-entropy volume scattering with low anisotropy
L10	Medium-entropy volume scattering with high anisotropy
L11	High-entropy volume scattering with low anisotropy
L12	High-entropy volume scattering with high anisotropy
L13	Low-entropy surface scattering with low anisotropy
L14	Low-entropy surface scattering with high anisotropy
L15	Medium-entropy surface scattering with low anisotropy
L16	Medium-entropy surface scattering with high anisotropy

being present in each resolution cell, and on A, showing the importance of the secondary scatterers.

IV. LATENT DIRICHLET ALLOCATION

A widely used statistical generative model for the discovery of hidden semantic structures behind collections of images is LDA [9]. Assuming each image $\mathbf{w}_d$, as a combination of N_d visual word tokens, $\mathbf{w}_d = \{w_{d1}, w_{d2}, \ldots, w_{dN_d}\}$, LDA discovers its latent semantics as a set of k topics, where the topics are distributions over a fixed dictionary of N_V visual words. They are supposed to reflect semantic categories. Therefore, each image containing different targets is represented as a mixture of the topics.

For estimating the required model parameters a and B, and inferring the posterior distributions (i.e., the topic distributions in the images), LDA uses approximate inference algorithms, such as variational expectation maximization [9].

Having the model parameters and the posterior distributions, LDA can generate every nth visual word token (w_{dn}) of each image $\mathbf{w}_d$ through the generative process

$$p(w_{\mathrm{dn}}|a, B) = \int p(\theta_d|a) \left[\sum_{z_{\mathrm{dn}}} p(z_{\mathrm{dn}}|\theta_d) p(w_{\mathrm{dn}}|z_{\mathrm{dn}}, B) \right] d\theta_d. \tag{4}$$

In this process, LDA chooses a k-dimensional Dirichlet random variable $\theta_d \sim \mathrm{Dir}(a)$, corresponding to $\mathbf{w}_d$, where a determines the Dirichlet distribution's prior [9]. Then, it selects a topic token z_{dn} from the topic mixture θ_d and picks w_{dn} from the multinomial probability distribution conditioned on the selected topic, $p(w_{\mathrm{dn}}|z_{\mathrm{dn}}, B)$. The matrix $B_{N_V \times k}$ parameterizes the visual word probabilities within the topics.

V. HIGH-LEVEL SEMANTICS DISCOVERY

Since the low-level semantics only refer to physical scattering properties, it is difficult to associate them with common land-cover categories. Therefore, in this section, we employ LDA to discover higher level semantics. To this end, we split each image into adjacent patches of 64×64 pixels and represent each image patch as a BoW by generating a histogram of the 16 labels in Table I assigned to that image patch; the selected patch size of 64×64 pixels resulted as a compromise

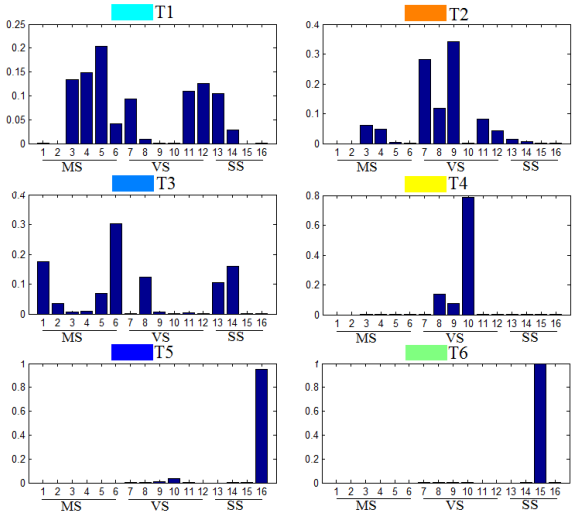

Fig. 4. Topics discovered by LDA. The x-axis denotes the 16 labels in Table I (e.g., $1 \rightarrow L1$), while the y-axis shows their probabilities.

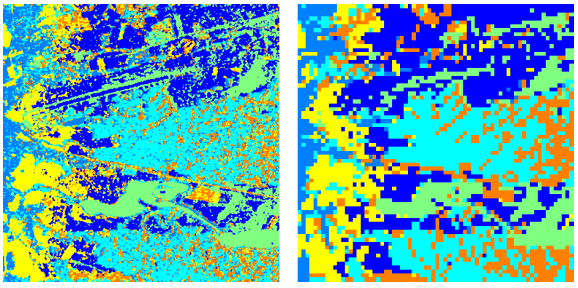

Fig. 5. Topic assignments to image pixels (left) and to image patches (right).

TABLE II
HIGH-LEVEL SEMANTICS

Topic	Semantics	Semantic rule
T1	Woody vegetation	53% MS + 34% VS + 13% SS
T2	Mixed woody vegetation & shrubs	12% MS + 86% VS + 02% SS
T3	Artificial, man-made structures	60% MS + 14% VS + 26% SS
T4	Herbaceous vegetation	00% MS +100%VS + 00% SS
T5	Smooth surface	00% MS + 05% VS + 95% SS
T6	Specular surface	00% MS + 00% VS + 100%SS

between small patches keeping the semantic analysis simple and bigger patches capturing the spatial context of objects. By considering each image patch as a document, we apply LDA to the BoW models to find latent semantics as a set of topics. In our experiments, the number of topics to be discovered was set to six based on an educated guess. The resulting topics are shown in Fig. 4 as probability distributions over the 16 labels. Fig. 5 shows the topic assignments to each image pixel and patch, where the assignments were made according to

$$\text{word } i \rightarrow \arg\{\max[p(w_i|z_k)]\} \tag{5}$$

$$\text{patch } j \rightarrow \arg\{\max[p(z_k|\theta_j)]\}. \tag{6}$$

By analyzing Figs. 4 and 5, and by using precisely coregistered Google Earth images as correct ground truth, it is possible to reveal the semantics of the topics and the rules connecting them to the lower-level semantics. As shown in Table II, the probabilities of the respective types of scattering can be written as percentages.

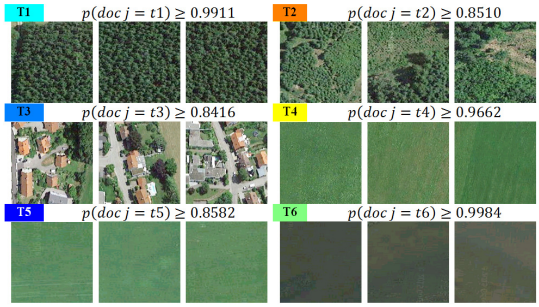

Fig. 6. Visualization of the discovered topic semantics. While discriminating the image patches of $T4$ and $T5$ is difficult due to their similar colors, they could be separated by using PolSAR images based on their different dominant scattering mechanisms (VS versus SS) caused by the distinct vegetation height.

1) *T1* is mostly assigned to forested regions. The high MS and VS values are due to the presence of trunks and rich canopies of woody vegetation. In addition, the low SS value may be caused by canopy tops or ground. According to Fig. 4, $T1$'s main component is $L5$, which is referred to in [7] as *double-bounce scattering in a high-entropy environment, such as vegetation, which has a well-developed branch and crown structure*. Therefore, we assign the semantic term *woody vegetation* to $T1$.
2) *T2* is most often assigned to regions covered by trees and shrubs. The very small SS value indicates the absence of rich canopies, which allows radar waves to scatter directly back to the sensor. The relatively large MS value shows the small number of tree trunks. $T2$'s main component is $L9$, which [7] refers to *vegetated surfaces with anisotropic scatterers and moderate correlation of scatterer orientations*. Thus, we assign the semantic term *mixed woody vegetation and shrubs* to $T2$.
3) *T3* is mostly characterized by MS caused by building walls and tree trunks. In addition, VS and SS in this topic can be caused by tree canopies and smooth surfaces (e.g., streets and roofs), respectively. $T3$ is the only topic with nonzero $L1$ and $L2$. These are referred to in [7] as *double-bounce scattering events provided by isolated dielectric and metallic dihedral scatterers*. Therefore, we assign the semantic term *artificial/man-made structures* to $T3$.
4) *T4* is characterized only by VS indicating that it usually occurs in regions covered by vegetation with thin stems, which do not backscatter the incident wave. Thus, we assign the semantic term *herbaceous vegetation* to $T4$.
5) *T5* is distinguished by a very large SS and a very small VS value. This reveals thin land-cover allowing radar waves to reach the ground. Therefore, we assign the semantic term *smooth surface* to $T5$.
6) *T6* is completely composed of SS. Therefore, we assign the semantic term *specular surface* to it.

Fig. 6 demonstrates the ground truth (Google Earth images) of some image patches, which are mainly characterized by a single topic, in order to visualize the semantics of each topic.

VI. Assessment of High-Level Semantics

In this section, we assess the discovered high-level semantics by modeling each image patch as a vector of the extracted topics (BoT model). These vectors form a multidimensional Euclidean space, the so-called topic space, in which each image patch is represented as a point positioned according to its topic distribution. Since the topic space dimensions are semantically meaningful (e.g., woody vegetation), the image patch distribution within this space is easily understandable. When these topics are the basic land-cover types, different topic mixtures result in various mixed land-cover types. Thus, by navigating through the topic space, users can easily reach the image patches containing their desired land-cover types and define new categories of mixed land-cover types.

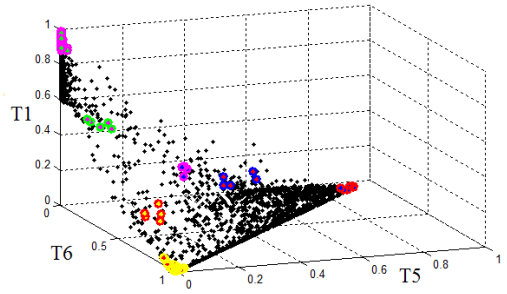

Fig. 7. Three-dimensional space spanned by $T5$, $T6$, and $T1$. The highlighted points are visualized in Fig. 8.

In order to visualize topic spaces, we select three topics and pick up the image patches containing them. After discarding the other topics, the chosen image patches are displayed in a 3-D normalized topic space. We explore the topic space and pick groupings of image patches from different regions and validate their semantics via their ground truth.

A. Natural Land-Cover Variation

Our aim is to discover different natural land-cover types as mixtures of the basic land-cover types: woody vegetation ($T1$), smooth land surfaces ($T5$), and specular surfaces ($T6$). Fig. 7 shows the topic space of these topics in which the coordinates are $(T1, T5, T6)$. Here, the selected image patches are highlighted and their ground truth is shown in Fig. 8.

In Fig. 7, starting from the point $(0, 0, 1)$, we expect to have image patches of specular surfaces, such as water bodies, which complies with Fig. 8 (grouping $G1$). By moving toward the point $(1, 0, 0)$, the proportion of the water bodies decreases while that of the forest increases, as demonstrated in Fig. 8 ($G2$) and ($G3$). Reaching the point $(1, 0, 0)$, the image patches are totally covered by forest, as shown in Fig. 8 ($G4$). Furthermore, around the point $(0.3, 0.3, 0.3)$, a mixture of the three basic land-cover types is expected, which is true according to Fig. 8 ($G5$). By moving toward the point $(0, 1, 0)$, the land-cover types change, containing higher proportions of grasslands, as shown in Fig. 8 ($G6$) and ($G7$).

B. Natural and Man-Made Land-Cover Variation

The idea is to discover various natural and man-made land-cover types as mixtures of the basic land-cover types: mixed woody vegetation and shrubs ($T2$), man-made structures ($T3$), and smooth surfaces ($T5$). Fig. 9 shows the topic space of the image patches in which the coordinates are $(T2, T3, T5)$.

As shown in Fig. 9, we selected three groupings of image patches from various regions: one grouping around

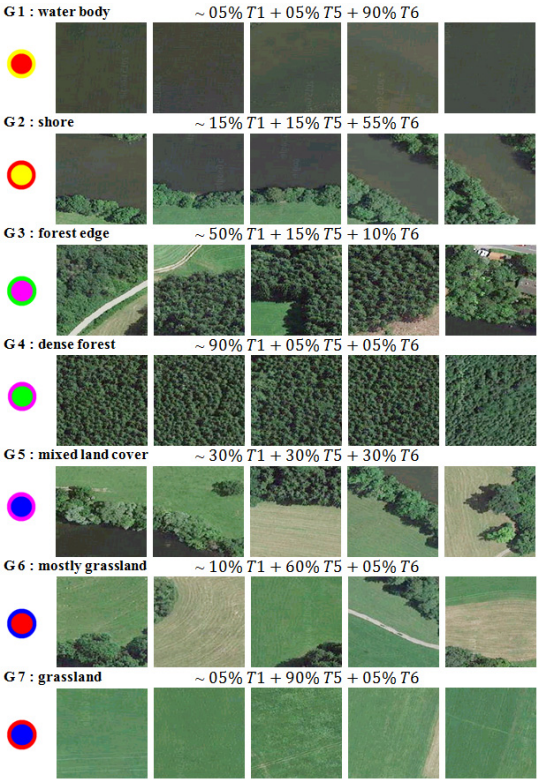

Fig. 8. Examples of new semantic classes spanned by $T5$, $T6$, and $T1$ corresponding to the highlighted points in Fig. 7.

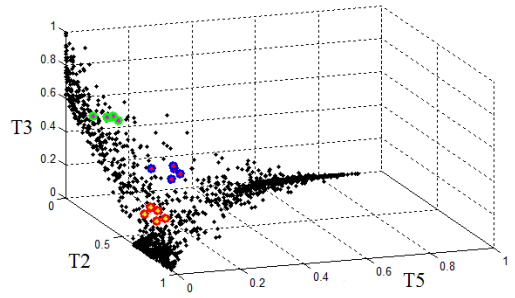

Fig. 9. Three-dimensional topic space spanned by $T5$, $T2$, and $T3$. The highlighted points are visualized in Fig. 10.

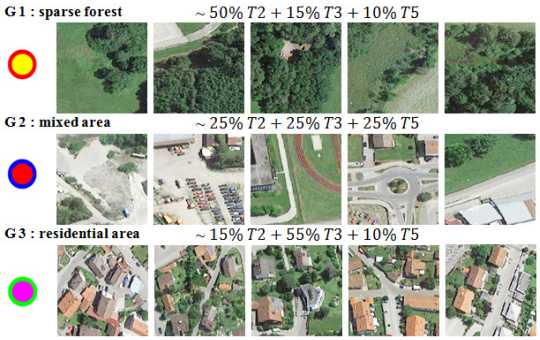

Fig. 10. Examples of new semantic classes spanned by $T5$, $T2$, and $T3$ corresponding to the highlighted points in Fig. 9.

the location $(0.5, 0.1, 0.1)$, one grouping around the location $(0.3, 0.3, 0.3)$, and one grouping around the location $(0.1, 0.6, 0.1)$. The visualization of their ground truth in Fig. 10 indicates that the image patches of the first grouping are mostly covered by sparse forest. The image patches of the second grouping contain mixtures of all the three basic land-cover types, while the third grouping mostly represents residential areas, including man-made structures (e.g., houses) and trees.

Altogether, representing images by their high-level semantics helps us to better understand their contents and discover new semantic categories as mixtures of high-level semantics. This further allows us to categorize images by selecting a set of coordinates and to partition the topic space.

VII. Conclusion

In this letter, we propose a multilevel approach for semantic relationship discovery in PolSAR images. First, we extract low-level semantics (physical properties of the recorded targets) by using the $H/A/\alpha - Wishart$ classification method. Then, the images are tiled into patches and modeled as BoWs by generating histograms of the class labels assigned to the image patches. Next, LDA is applied to the BoWs to discover higher level semantics as a set of topics. Our analysis shows that the topics refer to basic land-cover types and the image patches covered by other land-cover types (e.g., mixed land-cover types) are described as mixtures of the topics. Thus, modeling image patches by the topics (BoT) helps to better understand the existing semantic relationships and identify new land-cover type categories. The semantic relationships can be used in designing parameter-based land-cover categorization methods. Moreover, a topic representation of PolSAR images helps image mining systems to adapt their results to human semantics, thus narrowing the semantic gap.

References

[1] S. S. Durbha and R. L. King, "Semantics-enabled framework for knowledge discovery from Earth observation data archives," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 11, pp. 2563–2572, Nov. 2005.

[2] M. Datcu *et al.*, "Information mining in remote sensing image archives: System concepts," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2923–2936, Dec. 2003.

[3] K. Alonso and M. Datcu, "Accelerated probabilistic learning concept for mining heterogeneous Earth observation images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3356–3371, Jul. 2015.

[4] R. Bahmanyar, A. M. M. de Oca, and M. Datcu, "The semantic gap: An exploration of user and computer perspectives in Earth observation images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2046–2050, Oct. 2015.

[5] D. Bratasanu, I. Nedelcu, and M. Datcu, "Bridging the semantic gap for satellite image annotation and automatic mapping applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 193–204, Mar. 2011.

[6] S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 498–518, Mar. 1996.

[7] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, Jan. 1997.

[8] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2249–2258, Sep. 1999.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[10] R. Bahmanyar, S. Cui, and M. Datcu, "A comparative study of bag-of-words and bag-of-topics models of EO image patches," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1357–1361, Jun. 2015.

[11] R. Horn, A. Nottensteiner, A. Reigber, J. Fischer, and R. Scheiber, "F-SAR—DLR's new multifrequency polarimetric airborne SAR," in *Proc. IEEE IGARSS*, vol. 2. Jul. 2009, pp. 902–905.