Multimodal Probabilistic Latent Semantic Analysis for Sentinel-1 and Sentinel-2 Image Fusion

Ruben Fernandez-Beltran, Juan M. Haut, *Student Member, IEEE*, Mercedes E. Paoletti, *Student Member, IEEE*, Javier Plaza, *Senior Member, IEEE*, Antonio Plaza, *Fellow, IEEE*, and Filiberto Pla

***Abstract*— Probabilistic topic models have recently shown a great potential in the remote sensing image fusion field, which is particularly helpful in land-cover categorization tasks. This letter first studies the application of probabilistic latent semantic analysis (pLSA) and latent Dirichlet allocation to remote sensing synthetic aperture radar (SAR) and multispectral imaging (MSI) unsupervised land-cover categorization. Then, a novel pLSA-based image fusion approach is presented, which pursues to uncover multimodal feature patterns from SAR and MSI data in order to effectively fuse and categorize Sentinel-1 and Sentinel-2 remotely sensed data. Experiments conducted over two different data sets reveal the advantages of the proposed approach for unsupervised land-cover categorization tasks.**

***Index Terms*— Image fusion, land-cover categorization, probabilistic latent semantic analysis (pLSA), Sentinel-1, Sentinel-2.**

I. Introduction

THE recent online availability of Sentinel's operational products provides widespread opportunities to combine complementary information acquired by different sensors in order to conduct interdisciplinary research in many European Union policy-relevant application domains, such as land, marine, and atmosphere monitoring, climate change, and security services. Within the context of the Copernicus program, Sentinel-1 (S1) [1] and Sentinel-2 (S2) [2] missions exhibit a special synergy because their corresponding data products represent the Earth's surface in a fundamentally complementary way, using synthetic aperture radar (SAR) and high-resolution multispectral imaging (MSI). On the one hand, the S2 MSI instrument passively measures electromagnetic radiation that captures useful information on chemical properties of surfaces, such as nitrogen, carbon, or moisture. On the other hand, the S1 SAR sensor actively emits electromagnetic radiation to measure the returning signal and, consequently, the scattering characteristics of the objects in the scene. Whereas, MSI images are easy to interpret for the human visual system, their quality and availability can be strongly affected by adverse atmospheric conditions, which motivates the constant development of new techniques specifically designed for remote sensing [3]. In addition, SAR images can capture information through fog, smoke, rain, and clouds. However, their data applicability highly depends on the backscattering properties of the target surface, as well as on the presence of speckle. As a result, the complementary nature of both S1 and S2 instruments provides an excellent scenario to overcome the limitations of each individual sensor by means of an information fusion approach.

In the literature, different techniques have been successfully applied to fuse SAR and MSI data at three different integration levels [4]: decision level, pixel level, and feature level. In the decision-level approach, a separate predictor is initially estimated for each individual sensor and eventually a fused output is generated by combining all these independent results. It is the case of the work presented in [5], where a decision-based fusion model based on two independent aggregation levels was introduced. In particular, SAR and MSI data are initially preclassified using a support vector machine (SVM), and then an additional classifier is applied over these results to provide a global prediction using both SAR and MSI information.

Despite the effectiveness of the decision-based approach to combine data from different sources, the resulting performance for SAR and MSI data may become rather limited because the different modalities are independently analyzed and the fusion step eventually occurs as a postclassification process, which may be difficult to design. In this regard, pixel-based fusion methods pursue to directly combine several image pixels to derive a new fused image that contains enhanced spatial–spectral information. For example, Sukawattanavijit *et al.* [6] make use of a principal component analysis decomposition approach for fusing multiple image modalities at a pixel level, and finally, the data are labeled using a genetic algorithm together with an SVM classifier. Nonetheless, the general pixel-level fusion approach has shown to be not entirely suitable for SAR imagery because of the speckle noise typically

Manuscript received April 27, 2018; revised May 28, 2018; accepted June 1, 2018. Date of publication June 21, 2018; date of current version August 27, 2018. This work was supported in part by Generalitat Valenciana under Grant APOSTD/2017/007, in part by Spanish Ministry under Grant FPU14/02012-FPU15/02090 and Grant ESP2016-79503-C2-2-P), and in part by the Junta de Extremadura under Grant GR15005. *(Corresponding author: Ruben Fernandez-Beltran.)*

R. Fernandez-Beltran and F. Pla are with the Institute of New Imaging Technologies, University Jaume I, 12071 Castellón de la Plana, Spain (e-mail: rufernan@uji.es; pla@uji.es).

J. M. Haut, M. E. Paoletti, J. Plaza, and A. Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain (e-mail: juanmariohaut@unex.es; mpaoletti@unex.es; jplaza@unex.es; aplaza@unex.es).

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.2018.2843886

present in this sort of images and its high computational cost when dealing with large amounts of data.

Regarding the feature-based fusion level, these kinds of methods try to overcome some of the aforementioned limitations by combining attributes extracted from several sources in order to generate a data representation involving features of multiple sensors. That is, features extracted from SAR images can provide discriminatory object information to reduce some of the optical uncertainty that may occur in MSI imagery. For instance, Zhang *et al.* [7] study multiple kinds of features to effectively fuse remotely sensed optical and SAR data. Specifically, this letter considers four different optical features based on the gray-level cooccurrence matrix approach and polarimetric-based features extracted from SAR images.

Notwithstanding the effectiveness showed by all the aforementioned approaches under specific conditions, an alternative fusion research line has recently attracted the attention of the remote sensing research community. This approach takes advantage of the generative framework provided by probabilistic topic models [8]. In general, topic models represent a kind of generative statistical models that provide methods to express data as probability distributions according to their hidden semantic patterns instead of their low-level observable features. As a result, these kinds of models show a growing potential in remote sensing fusion tasks because they allow managing different data sources at a higher abstraction level. For instance, Zhong *et al.* [9] propose a multifeature fusion strategy that concatenates three complementary kinds of features, i.e., spectral, texture, and structural features, using topic models' characterizations to conduct remote sensing scene classification. However, this fusion scheme is still constrained by the use of topic models with a single modality because the remote sensing data fusion problem logically has a multimodal nature.

Indeed, it was not until recently that a multimodal topic model was successfully applied to fuse SAR and MSI data. Specifically, Bahmanyar *et al.* [10] presented a multisensor land-cover classification technique using a visual bag-of-words (vBoW) characterization scheme together with a multimodal variant of the latent Dirichlet allocation (LDA) model [11], which makes use of two different vocabularies to jointly represent SAR and MSI data modalities. Despite the potential of this recent LDA-based fusion approach to outperform individual single modality data, LDA is not the only type of topic model available in the literature and analyzing the effect of using different kinds of probabilistic topic models for fusing SAR and MSI remotely sensed data still remains an open-ended issue.

With the aforementioned considerations in mind, the contribution of this letter is focused on a twofold target. On the one hand, we study the application of the two main topic model families, i.e., the probabilistic latent semantic analysis (pLSA) [12] and LDA [11] models, within the remote sensing SAR and MSI unsupervised land-cover categorization fields. On the other hand, we introduce a novel pLSA-based fusion approach that pursues to uncover multimodal feature patterns from SAR and MSI data in order to effectively fuse Sentinel-1 and Sentinel-2 remote sensing imagery. The experimental part of the work conducted over two different data sets reveals the advantages of the proposed approach for unsupervised land-cover categorization tasks.

II. Background on Topic Models

Broadly speaking, topic models can be categorized within the families of two reference models, pLSA [12] and LDA [11]. Specifically, pLSA defines a semigenerative data model by introducing a latent context variable associated with the different word polysemy occurrences. The pLSA generative process is made as follows: 1) select a document d with probability $p(d)$; 2) pick a latent class z with probability $p(z|d)$; and 3) generate a word w with probability $p(w|z)$. However, this generative process is usually called ill-defined because documents set topic mixtures and simultaneously topics generate documents; thus, there is not a natural way to infer previously unseen documents [8]. In addition, the number of pLSA parameters grows linearly with the number of training documents, which makes this model particularly memory demanding and susceptible to overfitting [13].

In order to overcome pLSA limitations, Blei *et al.* [11] proposed the LDA model as a more general framework. In particular, LDA represents documents as a multinomial of topic mixtures generated by a Dirichlet prior that is able to predict new documents. Although LDA potentially overcomes pLSA limitations by means of using two Dirichlet distributions, one to model documents $Dir(\alpha)$ and another to model topics $Dir(\beta)$, the α and β hyperparameters have to be estimated during the topic extraction process, which logically adds an extra computational time and makes LDA performance highly sensitive to the quality of this initial estimation. In practice, α and β are estimated by iterating over the document collection that results in LDA requiring dense distributions to obtain a good hyperparameter estimation [14]. In fact, Chang *et al.* [15] stand that pLSA is able to obtain a topic structure more correlated to the human judgment than LDA, even though the perplexity metric may suggest the opposite.

All these facts make that pLSA-based models are usually preferred when few information is available according to the complexity of the data [16]. In the particular case of remotely sensed imagery, the high complexity of the visual patterns extracted from both SAR and MSI image domains generally makes that the amount of information available for a specific area of interest is rather limited to perform an effective unsupervised land-cover categorization. As a result, pLSA-based fusion models may take advantage of considering the document collection as model parameters to uncover more descriptive semantic patterns than LDA with less data. Precisely, this is the reason why the fusion method presented in this letter is based on pLSA.

III. Multimodal pLSA for Image Fusion

This section describes the three steps of the multimodal pLSA-based SAR and MSI fusion scheme presented in this letter (see Fig. 1): 1) image characterization; 2) multimodal pLSA-based image fusion; and 3) land-cover categorization.

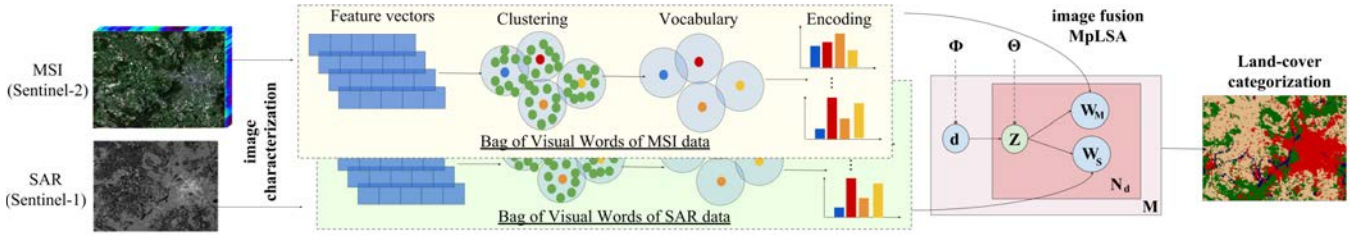

Fig. 1. Overview of the proposed fusion framework. In the image fusion step, we can observe the MpLSA graphical description, where d,z,w_S, and w_M represent the document, topic, SAR, and MSI word random variables. Besides, Φ and Θ represent the $p(z|d)$ and $p(w_S, w_M|z)$ model parameters, respectively.

A. Image Characterization

As image characterization scheme, we make use of the vBoW approach [17] that includes a three-step procedure. First, SAR and MSI coaligned data products are tiled into 32×32 image patches, which define topic model documents (d). Second, the k-means clustering algorithm [18] is globally applied over each image modality to build the corresponding SAR and MSI visual vocabularies. More specifically, we use vectorized 3×3 image patches with one pixel overlapping as local primitive features and a total number of 50 clusters, which represent the observable words for each modality, i.e., w_S for SAR and w_M for MSI. Finally, the local primitive features (vectorized 3×3 image patches) within each topic model document (a 32×32 image patch) are encoded in a single histogram of visual words by accumulating the number of local features into their closest clusters. Note that we use this straightforward feature description method for the sake of simplicity; however, other characterization approaches could be used instead. From this first image characterization step, we eventually generate a collection of M documents $D = \{d_1, d_2, \ldots, d_M\}$ characterized in both SAR and MSI visual vocabularies, i.e., $d_k = \{n(w_S^i, w_M^j, d_k)\} \forall i, j \in \{1, 2, \ldots, 50\}$, where $n(w_S^i, w_M^j, d_k)$ represents the number of times the SAR visual word w_S^i and the MSI term w_M^j cooccur within the document d_k.

B. Multimodal pLSA-Based Image Fusion

Based on the asymmetric pLSA formulation [12], we define a multimodal extension, called multimodal probabilistic latent semantic analysis (MpLSA), which is specially designed to fuse SAR and MSI data according to the aforementioned image characterization scheme. In particular, we extend pLSA by adding two diverging random variables to manage SAR and MSI modalities (see Fig. 1), that is, w_S represents SAR visual words and w_M expresses the MSI vocabulary.

In this letter, MpSLA parameters, i.e., $\Phi = p(z|d)$ (topic-document conditional probability distribution) and $\Theta = p(w_S, w_M|z)$ (multimodal word-topic distribution), are estimated by maximizing the complete log-likelihood function using the expectation-maximization (EM) algorithm that works in two stages: 1) E-step, where the expected value of the likelihood is computed given the current estimation of the parameters (1) and 2) M-step, where the new optimal values of the parameters are calculated according to the current setting of the hidden variables (2) and (3), as shown at the bottom of the next page

$$p(z|w_S, w_M, d) = \frac{p(w_S, w_M|z)p(z|d)}{\sum_z p(w_S, w_M|z)p(z|d)} \tag{1}$$

As convergence conditions for the MpLSA model, we use a 10^{-6} stability threshold in the difference of the log-likelihood between two consecutive iterations or a maximum number of 1000 EM iterations. Finally, we note that the Φ parameter provides the fused representation of the input data, which jointly models SAR and MSI features, and the Θ parameter contains the semantic hidden patterns of the multimodal data.

C. Land-Cover Categorization

Once the corresponding SAR and MSI data products have been fused according to the MpLSA model, we assume that each one of the K uncovered topics represents a land cover category. That is, the Θ parameter of MpLSA defines the semantic patters that we use to provide an Earth surface categorization based on the ground-truth information. In particular, each document is categorized according to the dominant topic, i.e., the highest probability value in Φ ($\text{argmax}_k\, p(z_k|d)$).

IV. Experiments

A. Data Sets

Two different Sentinel-1 (SAR) and Sentinel-2 (MSI) data products have been selected for the experiments.

1) Munich [10] includes a coupled Sentinel-1B (Level-1 ground range detected) and Sentinel-2A (Level-1C reflectance) data products of the City of Munich (Germany), acquired on September 29 and 30, 2016, respectively. They cover the Earth surface between the (48.33° N, 11.06° E) upper left coordinates and (47.77° N, 11.78° E) lower right coordinates. In the case of Sentinel-2, the B2, B3, B4, and B8 bands have been considered for the experiments because these bands have the highest spatial resolution, i.e., 10 m, and they also represent the blue, red, green, and infrared channels. Besides, Sentinel-1 data, which was initially acquired at 10.13-m spatial resolution, has been accordingly resampled to 10 m in order to obtain a final common size of 5596×6031 pixels.
2) Berlin [10] contains two Sentinel-1B and Sentinel-2A data products from Berlin (Germany), captured on

TABLE I
QUANTITATIVE ASSESSMENT OF THE UNSUPERVISED LAND-COVER CATEGORIZATION RESULTS FOR MUNICH AND BERLIN DATA SETS

	CATEGORY	MUNICH						BERLIN					
		SENTINEL-1 (SAR)		SENTINEL-2 (MSI)		FUSION (SAR+MSI)		SENTINEL-1 (SAR)		SENTINEL-2 (MSI)		FUSION (SAR+MSI)	
		LDA	pLSA	LDA	pLSA	MMLDA	MpLSA	LDA	pLSA	LDA	pLSA	MMLDA	MpLSA
PRECISION	Agriculture	78.49±0.17	80.81±0.03	83.98±0.12	84.54±0.16	84.56±0.03	84.69±0.09	94.38±0.04	87.66±0.03	74.57±0.62	81.61±7.37	87.5±0.6	90.64±0.07
	Forest	62.13±0.41	58.14±0.17	87.54±0.3	87.37±0.22	85.05±0.23	85.95±0.15	48.99±0.14	51.58±0.24	89.68±4.79	90.88±5.52	86.31±0.79	89.09±0.26
	Building	93.07±0.13	94.48±0.4	47.31±0.33	48.66±0.09	54.08±0.01	74.51±0.39	26.75±0.95	77.4±0.23	71.07±3.39	67.04±7.85	75.02±0.14	86.0±0.19
	Water	94.74±0.08	94.55±0.03	89.38±13.21	96.19±0.31	96.06±0.0	96.0±0.08	39.31±0.62	23.78±0.04	0.01±0.0	10.77±8.82	0.01±0.0	31.61±0.22
	AVG	**82.11±0.13**	81.99±0.15	77.05±5.61	**79.19±0.08**	79.94±0.09	**85.29±0.13**	52.36±0.37	**60.11±0.1**	58.83±1.96	**62.58±1.2**	62.21±0.32	**74.34±0.07**
RECALL	Agriculture	79.75±0.37	75.02±0.07	57.59±1.26	60.45±0.17	67.89±0.1	83.97±0.34	51.19±0.77	66.15±0.05	53.12±0.73	57.48±3.82	53.48±0.34	77.95±0.1
	Forest	84.33±0.26	87.43±0.07	89.87±0.36	90.15±0.15	89.6±0.02	89.99±0.1	94.03±0.13	96.56±0.02	74.67±0.37	70.18±7.98	89.57±0.53	92.27±0.11
	Building	32.94±0.72	27.83±0.9	77.74±0.46	77.39±0.25	73.36±0.37	70.35±0.07	17.95±0.41	15.31±0.72	69.77±2.77	80.67±1.43	72.82±1.76	82.9±0.25
	Water	87.74±0.06	87.78±0.11	93.28±0.49	93.07±0.49	92.31±0.04	92.39±0.04	82.97±0.25	82.58±0.15	0.08±0.03	58.66±47.88	0.07±0.0	92.69±0.14
	AVG	**71.19±0.24**	69.52±0.36	79.62±0.36	**80.26±0.13**	80.79±0.14	**84.18±0.12**	61.53±0.24	**65.15±0.28**	49.41±1.07	**66.75±18.97**	53.98±0.66	**86.45±0.06**
F-SCORE	Agriculture	79.11±0.1	77.81±0.04	68.32±0.92	70.49±0.07	75.32±0.07	84.33±0.13	66.37±0.63	75.4±0.04	62.04±0.7	67.43±5.03	66.38±0.31	83.82±0.05
	Forest	71.54±0.18	69.83±0.1	88.69±0.08	88.74±0.08	87.27±0.11	87.93±0.03	64.42±0.09	67.24±0.2	81.43±2.16	78.7±4.44	87.9±0.17	90.65±0.09
	Building	48.65±0.77	42.99±1.03	58.82±0.2	59.75±0.13	62.26±0.13	72.37±0.15	21.48±0.59	25.56±1.02	70.41±2.98	72.9±4.08	73.89±0.94	84.42±0.09
	Water	91.11±0.03	91.04±0.06	90.72±7.76	94.6±0.11	94.15±0.02	94.16±0.06	53.34±0.54	36.93±0.05	0.02±0.01	18.2±14.88	0.02±0.0	47.14±0.24
	AVG	**72.6±0.29**	70.42±0.42	76.64±3.2	**78.4±0.02**	79.75±0.04	**84.7±0.05**	**51.4±0.22**	51.28±0.4	53.48±1.17	**59.31±4.5**	57.05±0.36	**76.51±0.07**

May 26 and 27, 2017, respectively, which cover the area between the (52.78° N, 12.45° E) upper left coordinates and (52.26° N, 13.67° E) lower right coordinates. Again, both Sentinel-1 and Sentinel-2 data products have been processed following the aforementioned procedure in order to obtain a final size of 8149×5957.

Both data sets have been downloaded from the German Earth Observation Center website (http://goo.gl/ma9dUt) where ground-truth land-cover information (”Agriculture,” ”Building,” ”Forest,” and ”Water”) is also available for assessment purposes. Thus, the number of topics (K) has been fixed to four.

B. Results

Table I presents the quantitative evaluation of the unsupervised land-cover categorization results for tested data sets in terms of precision, recall, and f-score metrics. Specifically, ground truth image categories are shown in rows and Sentinel-1, Sentinel-2, and fusion results are provided in columns. It should also be noted that each table cell contains the average percentage and the corresponding standard deviation obtained after five runs of the indicated topic models. Also, Fig. 2 provides a qualitative assessment of the results by reporting the corresponding unsupervised land-cover categorization maps.

One of the first noticeable points that can be observed when comparing LDA and pLSA land-cover categorization effectiveness over S1 and S2 data is that, in the case of the Munich data set, pLSA obtains a better land-cover categorization result than LDA when considering MSI imagery. However, LDA seems to be more effective to deal with SAR data. In the case of Berlin, pLSA consistently provides a better metric result than LDA for both SAR and MSI data. These quantitative results are also supported by the corresponding land-cover categorization maps presented in Fig. 2, where pLSA shows a particularly relevant visual performance improvement over LDA for the Berlin MSI data. Even though both Munich and Berlin data sets logically have a similar nature, it is possible to observe that pLSA is able to work better with the higher complexity of the Berlin data set and especially with the MSI data. Note that some categories in the Berlin image are significantly unbalanced, which indicates that pLSA can take advantage of using the document collection as model parameters, whereas LDA's Dirichlet hyperparameter estimation may become rather inaccurate in this case. In fact, this is the reason why LDA performance drops dramatically in Berlin's ”Water” category.

Regarding the considered SAR and MSI multimodal data fusion schemes, both multimodal latent Dirichlet allocation (MMLDA) and MpLSA models generally provide a remarkable quantitative metric improvement with respect to their corresponding single-modal LDA and pLSA counterparts. Even though the general multimodal data fusion scheme logically helps to overcome each individual sensor limitations, the proposed MpLSA-based fusion model has proved to obtain an important advantage with respect to MMLDA for unsupervised land-cover categorization tasks. According to the results reported in Table I, MpLSA provides average precision, recall, and f-score improvements of 8.74, 17.93, and 12.20 percentage units, on average, over the two considered data sets. In addition, Fig. 2 shows that MpLSA obtains the most similar results to the corresponding ground-truth information.

In general, the SAR and MSI data fusion problem raises the challenge of simultaneously managing two different data modalities, which eventually leads to the need of dealing with more complex data. From the generative perspective

$$p(w_S, w_M|z) = \frac{\sum_d n(w_S, w_M, d)p(d)p(z|w_S, w_M, d)}{\sum_{w_S,w_M}\sum_d n(w_S, w_M, d)p(d)p(z|w_S, w_M, d)} \tag{2}$$

$$p(z|d) = \frac{\sum_w n(w_S, w_M, d)p(z|w_S, w_M, d)}{\sum_z\sum_{w_S,w_M} n(w_S, w_M, d)p(z|w_S, w_M, d)}. \tag{3}$$

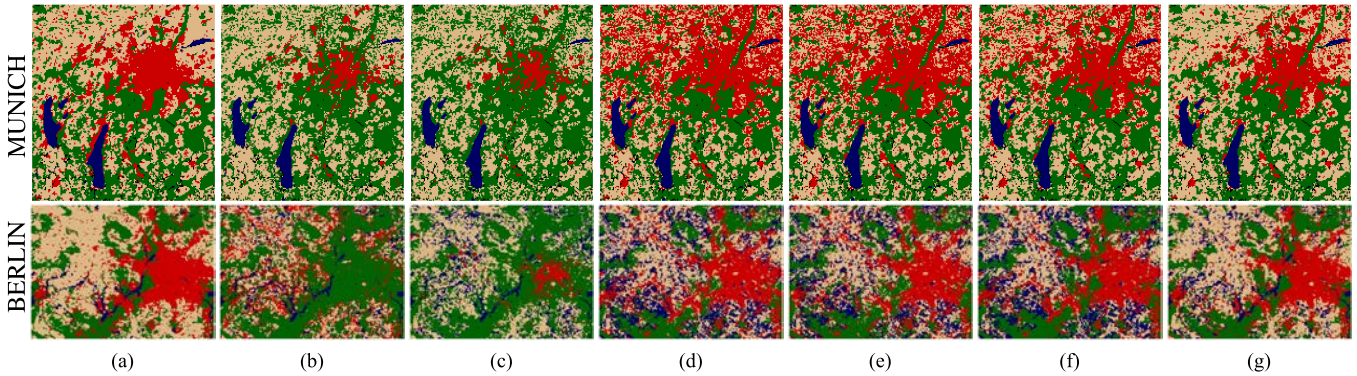

Fig. 2. Qualitative assessment of the unsupervised land-cover categorization results for Munich and Berlin data sets. (a) GT. (b) LDA-SAR. (c) pLSA-SAR. (d) LDA-MSI. (e) pLSA-MSI. (f) MMLDA. (g) MpLSA.

of topic models, the MMLDA-based fusion approach needs to estimate the α and β Dirichlet hyperparameters from the document collection. However, the multimodal estimation of these parameters may become rather inaccurate because the number of documents is constrained to the size of the interest area. In other words, the higher complexity of the multimodal SAR/MSI data makes that MMLDA may require more documents, i.e., a bigger region of interest, to estimate the Dirichlet hyperparameters under fair conditions with respect to the single-modality case. Nonetheless, the proposed MpLSA fusion scheme takes advantage of using input documents as model parameters because the whole document distribution is considered in the model's posterior computation, which allows MpLSA to uncover more descriptive multimodal patterns than MMLDA, despite the fact that the number of documents remains fixed.

V. Conclusion

This letter presents a multimodal pLSA-based SAR and MSI data fusion framework in order to effectively perform unsupervised land-cover categorization. Our experiments, conducted using two coupled Sentinel-1 (SAR) and Sentinel-2 (MSI) data products, reveal that the presented model provides a competitive advantage with respect to the multimodal LDA-based fusion scheme in terms of both quantitative and qualitative results. The main conclusion that arises from this letter is the MpLSA potential to fuse SAR and MSI data belonging to limited areas of interest where the amount of information may be rather constrained. In addition, single-modal pLSA has also shown to outperform LDA when dealing with undercomplete data. Future work will be aimed at extending the proposed model to deep fusion architectures.

Acknowledgment

The authors would like to thank Dr. R. Bahmanyar for his helpful assistance.

References

[1] M. Drusch *et al.*, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, May 2012.

[2] R. Torres *et al.*, "GMES Sentinel-1 mission," *Remote Sens. Environ.*, vol. 120, pp. 9–24, May 2012.

[3] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification," *ISPRS J. Photogramm. Remote Sens.*, to be published, doi: 10.1016/j.isprsjprs.2017.11.021.

[4] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.

[5] B. Waske and S. van der Linden, "Classifying multilevel imagery from SAR and optical sensors by decision fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1457–1466, May 2008.

[6] C. Sukawattanavijit, J. Chen, and H. Zhang, "GA-SVM algorithm for improving land-cover classification using SAR and optical remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 284–288, Mar. 2017.

[7] H. Zhang, H. Lin, and Y. Li, "Impacts of feature normalization on optical and SAR data fusion for land use/land cover classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1061–1065, May 2015.

[8] D. M. Blei, "Probabilistic topic models," *Commun. ACM*, vol. 55, no. 4, pp. 77–84, Apr. 2012.

[9] Y. Zhong, Q. Zhu, and L. Zhang, "Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6207–6222, Nov. 2015.

[10] R. Bahmanyar, D. Espinoza-Molina, and M. Datcu, "Multisensor Earth observation image classification based on a multimodal latent Dirichlet allocation model," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 459–463, Mar. 2018.

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

[12] T. Hofmann, "Probabilistic latent semantic analysis," in *Proc. 15th Conf. Uncertainty Artif. Intell.* San Mateo, CA, USA: Morgan Kaufmann, 1999, pp. 289–296.

[13] R. Fernandez-Beltran and F. Pla, "Incremental probabilistic latent semantic analysis for video retrieval," *Image Vis. Comput.*, vol. 38, pp. 1–12, Jun. 2015.

[14] R. Fernandez-Beltran and F. Pla, "Latent topics-based relevance feedback for video retrieval," *Pattern Recognit.*, vol. 51, pp. 72–84, Mar. 2016.

[15] J. Chang, S. Gerrish, C. Wang, J. L. Boyd-Graber, and D. M. Blei, "Reading tea leaves: How humans interpret topic models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 288–296.

[16] R. Fernandez-Beltran and F. Pla, "Prior-based probabilistic latent semantic analysis for multimedia retrieval," *Multimedia Tools Appl.*, to be published, doi: 10.1007/s11042-017-5247-z.

[17] Y. Zhang, R. Jin, and Z.-H. Zhou, "Understanding bag-of-words model: A statistical framework," *Int. J. Mach. Learn. Cybern.*, vol. 1, nos. 1–4, pp. 43–52, 2010.

[18] J. M. Haut, M. Paoletti, J. Plaza, and A. Plaza, "Cloud implementation of the K-means algorithm for hyperspectral image analysis," *J. Supercomput.*, vol. 73, no. 1, pp. 514–529, 2017.