Multimodal Classification of Remote Sensing Images

Edemir Ferreira Jr, Arnaldo de A. Araújo, and Jefersson A. dos Santos
Department of Computer Science, Universidade Federal de Minas Gerais (UFMG)
Av. Antônio Carlos, 6627 - Pampulha - Belo Horizonte - MG, CEP 31270-901, Brazil
edemirm, arnaldo, jefersson@dcc.ufmg.br

***Abstract*—Remote Sensing Images (RSIs) have been used as a major source of data, particularly with respect to the creation of thematic maps. This process is usually modeled as a supervised classification problem where the system needs to learn the patterns of interest provided by the user and assign a class to the rest of the image regions. Associated with the nature of RSIs, there are several challenges that can be highlighted: (1) they are georeferenced images, i.e., a geographic coordinate is associated with each pixel; (2) the data commonly captures specific frequencies across the electromagnetic spectrum instead of the visible spectrum, which requires the development of specific algorithms to describe patterns; (3) the detail level of each data may vary, resulting in images with different spatial and pixel resolution, but covering the same area; (4) due to the high pixel resolution images, efficient processing algorithms are desirable. Thus, it is very common to have images obtained from different sensors, which could improve the quality of thematic maps generated. However, this requires the creation of techniques to properly encode and combine the different properties of the images. Therefore, this M.Sc. dissertation[1] proposes two techniques for classification of regions in RSIs that manages to encode features extracted from different sources of data, spectral and spatial domains. The major objective is the development of approaches able to exploit the diversity of these different types of features to improve the accuracy in the creation of thematic maps.**

***Keywords*-Multimodal Classification; Remote Sensing; Data Fusion.**

I. Introduction

Over the years, there has been a growing demand for remotely-sensed data. Specific objects of interest are being monitored with earth observation data, for the most varied applications. Some examples include ecological science [1], hydrological science [2], agriculture [3], and many other applications.

RSIs have been used as a major source of data, particularly with respect to the creation of thematic maps. A thematic map is a type of map that displays the spatial distribution of an attribute that relates to a particular theme connected with a specific geographic area. This process is usually modeled as a supervised classification problem where the system needs to learn the patterns of interest provided by the user and assign a class to the rest of the image regions.

[1]This work relates to a M.Sc. dissertation.

In the last few decades, the technological evolution of sensors has provided remote sensing analysis with multiple and heterogeneous image sources, which can be available for the same geographical region: high spatial, multispectral, hyperspectral, radar, multi-temporal, and multiangular images can today be acquired over a given scene.

Typically, these sensors are designed to be specialists in obtaining one or few properties from the earth surface. This occurs because each sensor, due to technical and cost limitations, has a specific observation purpose and operates at different wavelength ranges to achieve it. Since the sensors are specialists, they carry different and complementary information, which can be combined to improve classification of the materials on the surface and consequently increase the quality of the thematic map. In this scenario, it is essential to use a more suitable technique to combine the different features in an effective way.

The remote sensing community has been very active in the last decade in proposing methods that combine different modalities [4]. In addition to support the research on this important topic, every year since 2006, the IEEE Geoscience Remote Sensing Society (GRSS) has been developing a Data Fusion Contest (DFC), organized by the Image Analysis and Data Fusion Technical Committee (IADFTC), which aims at promoting progress on fusion and analysis methodologies for multisource remote sensing data. Also, other data fusion challenges have been proposed more recently by the International Society for Photogrammetry and Remote Sensing (ISPRS), devoted to the development of international cooperation for the advancement of photogrammetry and remote sensing and their applications. All the effort to reach advance in this research area shows the high interest and timely relevance of the posed problems.

Multimodal classification is a challenging task for several reasons. First, the data are generated by very complex systems, driven by numerous underlying processes that depend on the sensor used and a large number of variables which sometimes we have no access, e.g., the atmospheric constituents cause wavelength-dependent absorption and scattering of radiation, which degrade the quality of images. Second, combining heterogeneous datasets such that the respective advantages of each dataset are maximally exploited, and drawbacks suppressed, is not an evident task. Third, as pointed by [5], it is very difficult to conclude what is the best approach for multimodal data fusion, since it depends on the foundation of the problem, the nature of the data used and the source of information utilized.

There are also several research challenges in computational scope when working with RSI classification such as: (1) remote sensing data is inherently big, even at 250 m coarse spatial resolution, Moderate-Resolution Imaging Spectroradiometer (MODIS) product can contain more than 20 millions of pixels, jointly with a time series of five thousand observations. Most machine learning models described as a state of the art (e.g., Deep Neural Networks, non-linear Support Vector Machines), can not handle with the magnitude of this data; (2) segmentation scale, accompanied by the large amount of information at the level of object in very high spatial resolution images, segmentation algorithms have difficulty in defining the optimum scale to be used; (3) pixel mixture and dimensionality reduction, images with high spectral resolution must be preprocessed due to problems such as high dimensionality, treatment of noise and corrupted bands, mixture of pixels due to the low spatial resolution; (4) efficiency, even collecting information from various sensors, efficiency and capability to process that amount of data is desired or even crucial depending on the application. In applications such as tsunami or earthquakes, the data must be analyzed in near real time, and the difference of a few seconds can save hundreds or even thousands of lives in a seaquake.

In this work, we are interested in the use of RSI particularly with respect to the creation of thematic maps by exploiting multi sensor data. To this purpose, we proposed two different approaches to the classification task, designed to receive two images, over the same geographic region, with different domains as input: an image with very high spatial (VHS) resolution and another one with multi/hyperspectral (HS) resolution (Figure 1).

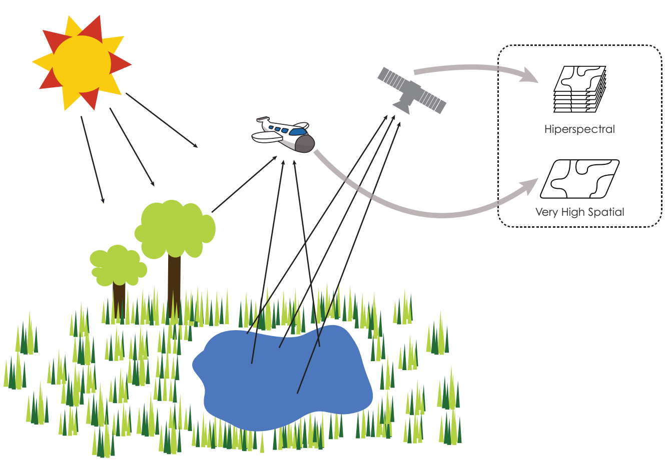

Fig. 1. An illustration of multimodal data acquisition. The figure shows two different platforms: a plane and a satellite; carrying sensors which extract different information (spectral and spatial) over the same region, creating a multimodal perspective.

The first approach is called Dynamic Majority Vote. In this approach, we created a framework based on a supervised learning scheme, divided in six steps: (1) data acquisition, the framework receives the VHS and HS images, acquired by different sensors but over the same geographic area as input; (2) object representation, the VHS image is segmented in regions using a segmentation algorithm while the HS image is analyzed by the spectral signature of each pixel; (3) feature extraction, feature vectors are extracted from the segmented regions of VHS using various descriptors and the spectral signatures are obtained by different dimensionality reduction methods; (4) training, using different learning methods and the feature vectors extracted from both domains a set of base classifiers is created; (5) dynamic weight matrix construction, using the trained base classifiers and a validation set is create a matrix of weights which represents the importance of each classifier at decision in every class; (6) prediction, given unseen samples and built the dynamic weight matrix, a predict for every new sample is made using the weights of the decisions of every classifier at that sample. Our approach has the ability to exploit these classifiers which have a specialty in some specifics classes, but would be suppressed by the other classifiers in an equal weight scheme.

The second approach is a boosting-based approach based on the SAMME Adaboost [6]. Such as the Dynamic Majority Vote, the boosting-based approach uses a supervised learning framework, but may be divided in five steps: (1-3) data acquisition, object representation and feature extraction as described in the previous description; (4) training, using the features extracted from both domains and diverse learning methods a set of weak learners is created, which at every boosting iteration once is selected to compose the final strong classifier; (5) prediction, given the unseen samples and the set of selected weak learners, a predict for every new sample is made regarding to the linear combination of the weak learner predictions. In this approach, we exploit the inherent feature selection of the Adaboost for the combination of different modalities, as a natural process.

To summarize, this work has the following two main contributions:

- A late fusion technique, called Dynamic Majority Vote, which exploits the specialty of different classifiers and combines them for a final decision for each pixel in the thematic map;
- A boosting-based approach, capable of combining different modalities of sensor data by using the inherent feature selection of the boosting-based strategy.

II. Related work

In data fusion, each data source describing the same scene and objects of interest can be defined as a *modality*. In remote sensing image analysis, the different modalities often represent a particular data property carrying complementary information about the surface observed [7].

The joint complementarity exploitation of different remote sensing sources has proven to be very useful in many applications of land-cover classification, and the capability of improving the discrimination between the classes is a key aspect towards a detailed characterization of the earth [5]. Concerning multisource data, a diversity of fusion techniques has been proposed in the remote sensing literature, which can be divided into levels according to the modalities used in the fusion, as follows:

1) **Fusion at subpixel level**: Given k modalities datasets, which usually involve different spatial scales, the modalities are fused at subpixel level using appropriate transforms [8]. These fusions are commonly used in the cases where the main objective is to preserve the valuable spectral information from multispectral or hyperspectral sensors, with low spatial resolution, as an alternative to pan-sharpening methods which can produce a spectral distortion [9].
   In the subject of proposed works based on spectral unmixing for data fusion, the spatial and temporal adaptive reflectance fusion model proposed in [10], was used in [11] for combining information from Landsat (30-m resolution) and MODIS (250-m to 1-km resolution), and a set of methods for increasing spatial resolution associated to [12] was used for classification task [13], [14]. An overview of the majority of nonlinear unmixing methods used in hyperspectral image processing and many recent developments in remote sensing are presented with details in [[15], [16]].
2) **Fusion at pixel level**: Given k modalities datasets, in the fusion at pixel level exists a direct pixel correlation between the modalities, which is used to produce data fusion. In general, that fusion level attempts to combine data from different sources in intent to produce a new modality, which, afterward, could be used for different applications. Some examples that rely on that case is pan-sharpening, super resolution, and 3D reconstruction from 2D views [5]. An evaluation of spatial and spectral effectiveness of more common pixel-level fusion methods was realized in [17]. Regarding [17], several pan sharpening methods have been proposed in the literature [[18], [19], [20]], primarily based on algebraic operations, component substitution, high-pass filtering and multi resolution analysis.
   More recently, [21] made an analysis of the different fusion techniques in images, also applied to remote sensing at a pixel level, showing that all techniques have their own limitation when used individually and they also encouraged the utilization of hybrid systems.
3) **Fusion at feature level**: Given k modalities datasets, various features are extracted individually from each modality, e.g., edges, corners, lines, texture parameters, followed by a fusion, which involves extraction and selection of more discriminant attributes. Regarding [4], one of the new research directions on feature level multimodal fusion are the Kernel methods. At the domain of remote sensing, there is a considerable number of studies about kernel methods [22], once they provide an instinctive way to encode data from different modalities into classification and prediction models. One of the first attempts to combine data from different modalities, using a combination of kernel functions, was realized by [23], who created a compound kernel by using the weighted summation of spatial and spectral features from the co-registered region. Extending the proposition for more than two sources, a multiple kernel learning [24] was applied to [25] for combining spatial and spectral information, to combine optical and radar data [26], [27], using the same sensor but in different places [28], also using different optical sensors to change detection [29].
4) **Fusion at decision level**: Given k modalities datasets, an individual process path is made for each modality, followed by a fusion of the outputs, assuming that the k outputs combined can improve the final accuracy [30]. In this way, the combination of complementary information from different modalities is done through the fusion of the results obtained considering each modality independently. There are several ways to combine the decisions, such as including voting methods, statistical methods, fuzzy logic-based methods, etc. When the results are explained as confidences instead of decisions, the methods are called soft fusion; otherwise, they are called hard fusion. An example of this type of fusion was presented in the 2008 [31] and 2009-10 [32] data fusion contests. [33] used a scheme of weighted decision fusion, which uses the SVM and the Random Forest for the probability estimation in the Landsat 8 and MODIS sensors; [34] made a combination of fusion by feature level using a graph-based feature fusion method together with a weight majority voting of outputs from differents SVM's for the classification of hyperspectral and Light Detection and Ranging (LiDAR) data.

The above-described levels do not cover all the possible fusion methods since input and output of data fusion can be different for each level of processing. In most cases, the fusion procedure is a junction of the four fusion levels considered previously.

The following two sections detail both contributions obtained.

III. Dynamic Majority Vote

We created a framework based on a supervised learning scheme, dealing with different scenarios, regions and objects, on the creation of thematic maps for the classification task. For that, we proposed a new approach, at decision level, to handle an amount of decisions from different classifiers, and combine them to obtain a final decision for each pixel in the thematic map. Contrary to approaches from the literature, our method uses the kappa index [35] to compare two classifiers. This fact brings some advantages since kappa index is more robust in dealing with unbalanced training sets.

The method is projected to receive two images from the same place with different domains as input: an image with very high spatial (VHS) resolution and another one with hyperspectral (HS) resolution. Our method is developed for a multiclass mapping scenario. It exploits the expertise of each learning approach over each class in order to find the most specialized classifiers. The result of this process is a *dynamic weight matrix*.

Most voting methods use an unique weight assigned to each classifier, regardless of the class to be predicted. This approach does not exploit the specialty of each classifier in a particular class, and thus can weaken the final model with no reliable predictions. Another weakness of the traditional majority voting is the difficulty of dealing with classifiers that produce similar mistakes in their predictions, thus resulting in the prediction of incorrect class.

The proposed approach uses a method for assigning weights where each classifier has a degree of reliability for each class to be predicted, resulting in dynamic weights. In addition, the method was created to handle classifiers that produce similar mistakes in their predictions. For this, an update of weights is accomplished favoring models whose distribution errors is uniform, thus hindering the allocation of high weight for classifiers not experts in difficult samples.

Our approach is divided into five main steps: object representation, feature extraction, training, dynamic weight matrix construction, and predicting. Figure 2 illustrates the proposed framework.

The contribution published in [36] evaluated the Dynamic Majority Vote in an urban scenario conducting a series of experiments in the IEEE GRSS Data Fusion Contest 2014 dataset, that demonstrated a significant improvement using the Kappa index and Overall Accuracy metrics in comparison with the proposed baselines. Our approach extracted features from different domains, which were trained with different learning techniques. This process created a set of classifiers with different expertise. The method assigned a weigh for each classifier according to their expertise in each specific class. The creation of the final thematic maps consisted in classifying each non-labeled region by fusioning the predicted output of each classifier according to their weights. This method was created by observing the latest approaches in multimodal classification in literature. The mostly uses the majority voting system as late fusion technique not exploiting the specialty of the classifiers in particular classes. So exploring the weakness of the traditional majority vote to deal with classifiers that produce similar mistakes in their predictions, we created a technique which prevents the allocation of high weight for classifiers not experts in difficult samples.

A more detailed discussion about the evaluation of the method and the results can be found in [36].

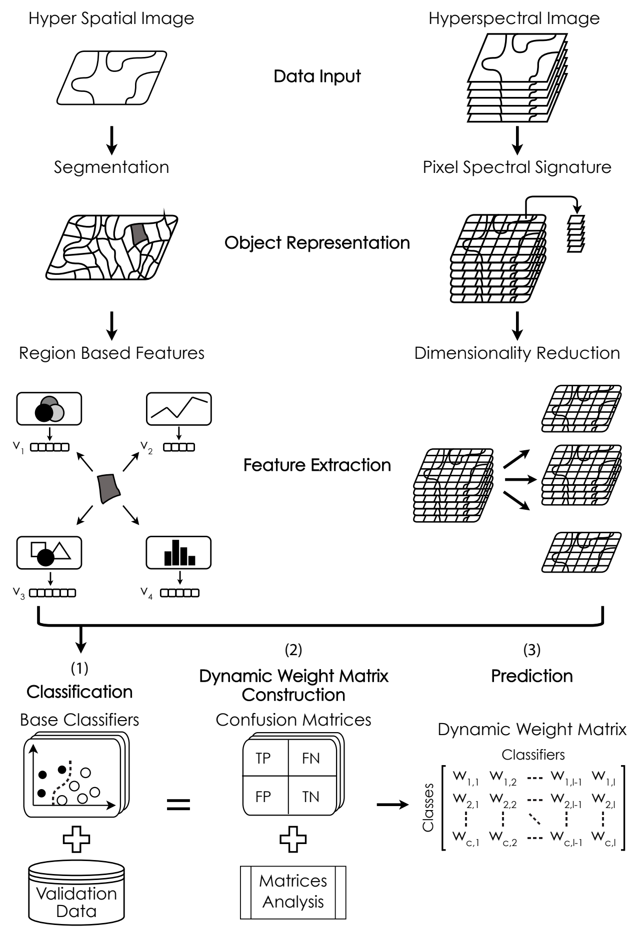

Fig. 2. The proposed *Dynamic Weight Matrix* (DWM)-based framework. Given the amount of feature extracted from both domains, a set of classifiers is create in (1). Afterwards, the classifiers are used in a validation data, producing a collection of confusion matrices, one for each classifier. Since built the collection of confusion matrices, an analyze of the distribution errors is made in (2) and the dynamic weight for every classifier in each class is produced. At the end, built the dynamic weight matrix and given the output of the classifiers in an unseen sample, the reliability of each prediction is consulted and the class with the highest final weight is chosen.

IV. Boosting-Based Approach

In this method, we aimed at exploiting multi-sensor data in a more general way, using the idea of boosting of classifiers, based on the SAMME Adaboost method [6].

The choice of an approach based on boosting is related to the inherent advantages of the strategy and its application in a multimodal classification of RSIs. Regarding the advantages, we can highlight: (1) algorithm flexibility, being possible to combine any method of learning as well extracted features obtained from different domains; (2) efficiency, when dealing with RSIs, the use of robust and efficient methods is desired, due to the complexity of the data (e.g., images with hundreds of spectral bands, very high pixel and spatial resolutions) and the high computational cost for processing; (3) tuning parameters, unlike most of the robust methods in the literature (e.g., SVMs, Neural Networks) that use non-linear models thus requiring various parameter settings, the boosting approach uses a combination of weak linear models to create a more complex function, and requires only a single parameter, the number of rounds to be trained; (4) well-known algorithm, in addition to the solid mathematical foundation behind the method, the literature also indicates successful works using boosting in remote sensing [37] and for other applications to computer vision [38].

We created a framework based on a supervised learning scheme, dealing with different scenarios, regions, and objects, on the creation of thematic maps for the classification task. We proposed a scheme, with a combination of pixel, feature, and decision levels, to handle an amount of information from different modalities, and combine them for a final decision for each pixel in the thematic map. Contrary to approaches from the literature, our method uses the inherent feature selection of the boosting for the combination of different modalities, as a natural process.

The proposed method is projected to receive two images from the same place with different domains as input: an image with very high spatial resolution and another one with hyperspectral resolution.

The boosting approach is divided into five main steps: data input, object representation, feature extraction, training, predicting. Figure 3 illustrates the proposed framework.

In the contribution published in [39], we evaluated the Boosting-based approach in an urban scenario and coffee crop recognition conducting a series of experiments in two datasets referred in urban multi-class scenario, and coffee crop recognition. The experiments demonstrated a significant improvement using the Kappa index and Overall Accuracy metrics in comparison with the proposed baselines at the urban scenario, but not statistically relevant in concern to the coffee crop recognition dataset using the Overall Accuracy. In that case, the main problem was the great difference between the pixel resolution of the images and a poor spectral information from a multi-spectral image, making unfeasible the extraction and combination of spectral information effectively.

Please refer to [39] for a more detailed discussion.

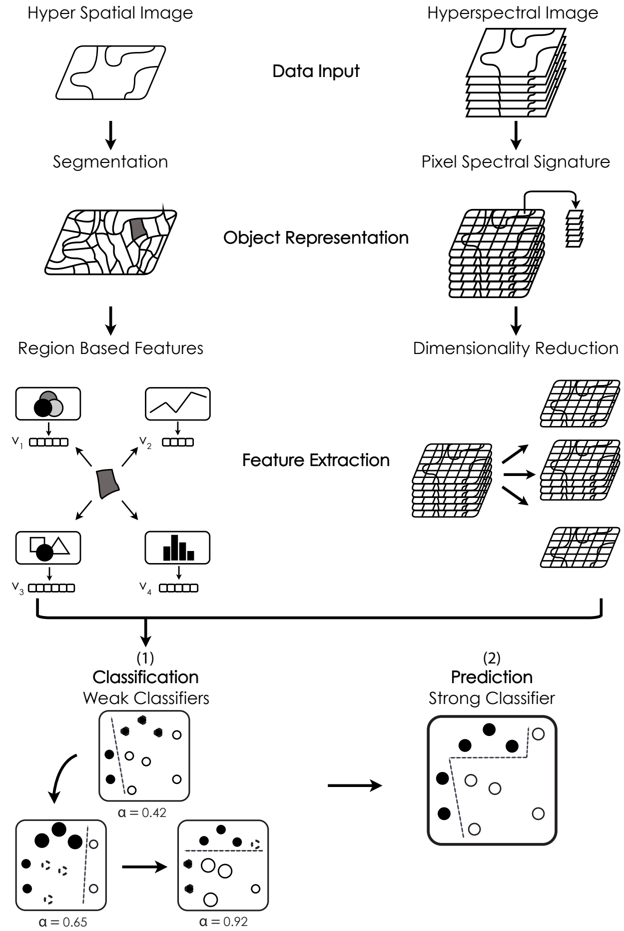

Fig. 3. The proposed *Boosting-based approach* framework. (1) The proposed method is projected to receive two images from the same place with different domains as input: an image with very high spatial resolution and another one with hyperspectral resolution; (2) the VHS image is segmented into regions using a segmentation algorithm while the HS image is analyzed by the spectral signature of each pixel; (3) feature vectors are extracted from the segmented regions of VHS using various descriptors and the spectral signatures are projected by using different dimensionality reduction methods; Given the amount of feature extracted from both domains, the boosting training starts in (4), where for every round one weak classifier will be chosen to compose the final strong classifier. The samples which are incorrectly labeled in every round, have their weight increased and will be focused by the learners in the next round. The collection of selected weak classifiers are combined in (5) to build the strong final classifier, which is used to predict the samples of the test data regarding the confidence of each weak model.

V. Conclusion

In this M.Sc. dissertation we addressed the use of RSI particularly with respect to the creation of thematic maps exploiting multi-sensor data. We dealt with two main challenges: the combination of referenced images from different domains (spatial and spectral) and how to exploit different types of features, extracted from these sensors.

To this purpose, we proposed two different approaches to the classification task, projected to receive two images, over the same geographic region: Dynamic Majority Vote and Boosting-based approach.

The joint complementarity exploitation of different remote sensing sources has proven to be very fruitful in the urban scenario dataset proposed, however the efforts to combine the information from a multispectral data with great difference of spatial resolutions and a poor spectral resolution prevent the Boosting-based approach to utilize the spectral features as additional information for the strong model.

This M.Sc. dissertation work was completed in two years (from March 2014 to March 2016) and has resulted in two conference papers [36] and [39].

As future work, we plan to evaluate the Boosting-based approach using other remote sensing datasets which contain spatial and hyperspectral information, and covering a higher spatial area. As another future work, we intend to adapt the framework to handle with different sensors, e.g., LIDAR, which contain elevation information from the objects.

Acknowledgment

This work was partially financed by CNPq (grant 449638/2014-6), CAPES, and Fapemig (APQ-00768-14). We are also grateful to Cooxupé and Rubens Lamparelli due to the support related to agricultural aspects and the remote sensing dataset.

References

[1] A. Ghiyamat and H. Z. Shafri, "A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment," *International Journal of Remote Sensing*, vol. 31, no. 7, pp. 1837–1856, 2010.

[2] T. Schmid, M. Koch, and J. Gumuzzio, "Multisensor approach to determine changes of wetland characteristics in semiarid environments (central spain)," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 11, pp. 2516–2525, 2005.

[3] Y. Lanthier, A. Bannari, D. Haboudane, J. R. Miller, and N. Tremblay, "Hyperspectral data segmentation and classification in precision agriculture: A multi-scale analysis," *Geoscience and Remote Sensing Symposium, IEEE International*, vol. 2, pp. II–585, 2008.

[4] L. Gomez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, "Multimodal classification of remote sensing images: A review and future directions," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 103, no. 9, pp. 1560–1584, 2015.

[5] M. D. Mura, S. Prasad, F. Pacifici, P. Gamba, and J. Chanussot, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1585–1601, 2015.

[6] J. Zhu, H. Zou, S. Rosset, and T. Hastie, "Multi-class adaboost," *Statistics and its Interface*, vol. 2, no. 3, pp. 349–360, 2009.

[7] I. R. Farah, W. Boulila, K. S. Ettabaa, and M. B. Ahmed, "Multiapproach system based on fusion of multispectral images for land-cover classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 12, pp. 4153–4161, 2008.

[8] S. Delalieux, P. J. Zarco-Tejada, L. Tits, M. A. Jimenez Bello, D. S. Intrigliolo, and B. Somers, "Unmixing-based fusion of hyperspatial and hyperspectral airborne imagery for early detection of vegetation stress," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 7, no. 6, pp. 2571–2582, 2014.

[9] B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu, "Spatial and spectral image fusion using sparse matrix factorization," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 3, pp. 1693–1704, 2014.

[10] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the landsat and modis surface reflectance: Predicting daily landsat surface reflectance," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, no. 8, pp. 2207–2218, 2006.

[11] C. M. Gevaert and F. J. García-Haro, "A comparison of starfm and an unmixing-based algorithm for landsat and modis data fusion," *Remote Sensing of Environment*, vol. 156, pp. 34–44, 2015.

[12] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 37, no. 3, pp. 1212–1226, 1999.

[13] R. Zurita-Milla, J. G. Clevers, and M. E. Schaepman, "Unmixing-based landsat tm and meris fr data fusion," *Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 453–457, 2008.

[14] J. Amorós-López, L. Gómez-Chova, L. Alonso, L. Guanter, J. Moreno, and G. Camps-Valls, "Regularized multiresolution spatial unmixing for envisat/meris and landsat/tm image fusion," *Geoscience and Remote Sensing Letters*, vol. 8, no. 5, pp. 844–848, 2011.

[15] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 7, no. 6, pp. 1844–1868, 2014.

[16] C. Lanaras, E. Baltsavias, and K. Schindler, "Advances in hyperspectral and multispectral image fusion and spectral unmixing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 40, no. 3, p. 451, 2015.

[17] J. Marcello, A. Medina, and F. Eugenio, "Evaluation of spatial and spectral effectiveness of pixel-level fusion techniques," *Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 432–436, 2013.

[18] J. Zhang, "Multi-source remote sensing data fusion: status and trends," *International Journal of Image and Data Fusion*, vol. 1, no. 1, pp. 5–24, 2010.

[19] I. Amro, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "A survey of classical methods and new trends in pansharpening of multispectral images." *Journal on Advances in Signal Processing*, vol. 2011, p. 79, 2011.

[20] T. Stathaki, *Image fusion: algorithms and applications*. Academic Press, 2011.

[21] R. Gharbia, A. T. Azar, A. E. Baz, and A. E. Hassanien, "Image fusion techniques in remote sensing," *arXiv preprint arXiv:1403.5473*, 2014.

[22] G. Camps-Valls and L. Bruzzone, *Kernel methods for remote sensing data analysis*. John Wiley & Sons, 2009.

[23] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93–97, 2006.

[24] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "Simplemkl," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.

[25] D. Tuia, F. Ratle, A. Pozdnoukhov, and G. Camps-Valls, "Multisource composite kernels for urban-image classification," *Geoscience and Remote Sensing Letters*, vol. 7, no. 1, pp. 88–92, 2010.

[26] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. L. Rojo-Álvarez, and M. Martínez-Ramón, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 6, pp. 1822–1835, 2008.

[27] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 10, pp. 3780–3791, 2010.

[28] L. Gómez-Chova, G. C. Valls, L. Bruzzone, and J. C. Maravilla, "Mean map kernel methods for semisupervised cloud classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 1, pp. 207–220, 2010.

[29] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 107, pp. 50–63, 2015.

[30] W. Li, S. Prasad, and J. E. Fowler, "Decision fusion in kernel-induced spaces for hyperspectral image classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 6, pp. 3399–3411, 2014.

[31] G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, C. Thiel, J. Inglada, E. Christophe, J. Chanussot *et al.*, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 grs-s data fusion contest," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 11, pp. 3857–3865, 2009.

[32] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot *et al.*, "Multi-modal change detection, application to the detection of flooded areas: outcome of the 2009–2010 data fusion contest," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 1, pp. 331–342, 2012.

[33] J. Wang, C. Li, and P. Gong, "Adaptively weighted decision fusion in 30 m land-cover mapping with landsat and modis data," *International Journal of Remote Sensing*, vol. 36, no. 14, pp. 3659–3674, 2015.

[34] W. Liao, R. Bellens, A. Pizurica, S. Gautama, and W. Philips, "Combining feature fusion and decision fusion for classification of hyperspectral and lidar data," *Geoscience and Remote Sensing Symposium, IEEE International*, pp. 1241–1244, 2014.

[35] R. G. Congalton and K. Green, *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press, 2008.

[36] E. F. de Andrade Jr, A. de Albuquerque Araújo, and J. A. dos Santos, "A multiclass approach for land-cover mapping by using multiple data sensors," *Iberoamerican Congress on Pattern Recognition*, pp. 59–66, 2015.

[37] J. A. dos Santos, P.-H. Gosselin, S. Philipp-Foliguet, R. d. S. Torres, and A. X. Falcao, "Multiscale classification of remote sensing images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 10, pp. 3764–3775, 2012.

[38] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Computer Vision and Pattern Recognition*, vol. 1, pp. I–511, 2001.

[39] E. F. de Andrade Jr, A. de Albuquerque Araújo, and J. A. dos Santos, "A boosting-based approach for remote sensing multimodal image classification," *SIBGRAPI 2016 Conference on Graphics, Patterns and Images, São José dos Campos, Brazil*, 2016.