HAL
open science

Challenges and Opportunities of Multimodality and Data Fusion in Remote Sensing

Mauro Dalla Mura, S Prasad, F Pacifici, P Gamba, Jocelyn Chanussot, Jon Atli Benediktsson

To cite this version:

Mauro Dalla Mura, S Prasad, F Pacifici, P Gamba, Jocelyn Chanussot, et al.. Challenges and Opportunities of Multimodality and Data Fusion in Remote Sensing. Proceedings of the IEEE, 2015, 103 (9), pp.1585-1601. 10.1109/JPROC.2015.2462751 . hal-01259771

HAL Id: hal-01259771

https://hal.science/hal-01259771v1

Submitted on 24 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Challenges and opportunities of multimodality and data fusion in remote sensing

M. Dalla Mura, *Member, IEEE,* S. Prasad, *Senior Member, IEEE,* F. Pacifici, *Senior Member, IEEE,* P. Gamba, *Fellow, IEEE,* J. Chanussot, *Fellow, IEEE,* J. A. Benediktsson, *Fellow, IEEE*

***Abstract*—Remote sensing is one of the most common ways to extract relevant information about the Earth and our environment. Remote sensing acquisitions can be done by both active (synthetic aperture radar, LiDAR) and passive (optical and thermal range, multispectral and hyperspectral) devices. According to the sensor, a variety of information about the Earth's surface can be obtained.**

The data acquired by these sensors can provide information about the structure (optical, synthetic aperture radar), elevation (LiDAR) and material content (multi and hyperspectral) of the objects in the image. Once considered together their complementarity can be helpful for characterizing land use (urban analysis, precision agriculture), damage detection (e.g., in natural disasters such as floods, hurricanes, earthquakes, oil-spills in seas), and give insights to potential exploitation of resources (oil fields, minerals). In addition, repeated acquisitions of a scene at different times allows one to monitor natural resources and environmental variables (vegetation phenology, snow cover), anthropological effects (urban sprawl, deforestation), climate changes (desertification, coastal erosion) among others. In this paper, we sketch the current opportunities and challenges related to the exploitation of multimodal data for Earth observation. This is done by leveraging the outcomes of the Data Fusion contests, organized by the IEEE Geoscience and Remote Sensing Society since 2006. We will report on the outcomes of these contests, presenting the multimodal sets of data made available to the community each year, the targeted applications and an analysis of the submitted methods and results: How was multimodality considered and integrated in the processing chain? What were the improvements/new opportunities offered by the fusion? What were the objectives to be addressed and the reported solutions? And from this, what will be the next challenges?

***Index Terms*—Data fusion, remote sensing, classification, pan-sharpening, change detection.**

I. Introduction

REMOTE sensing technologies can be used for observing different aspects of the Earth's surface, such as the spatial organization of objects in a particular region, their height, identification of the constituent materials, characteristics of the material surfaces, composition of the underground, etc. Typically, a remote sensing acquisition can just observe one (or few, at the most) of the aforementioned characteristics. Thus, the observations derived by different acquisition sources can be coupled and jointly analyzed by *data fusion* (DF) practices to achieve a richer description of the scene. The joint exploitation of different remote sensing sources is therefore a key aspect towards a detailed and precise characterization of the Earth. Fusion of multi-source information is nowadays considered to be a typical scenario in the exploitation of remote sensing data. Passive optical sensors have been widely used to map horizontal structures like land cover types at large scales, whereas Synthetic Aperture Radar (SAR) systems complement the optical imaging capabilities because of the constraints on time-of-day and atmospheric conditions and because of the unique responses of terrain and man-made targets to radar frequencies. Lately, Light Detection And Ranging (LiDAR) technology has proven to be uniquely positioned to provide highly accurate sample measurements of vertical height of structures (measure correlated to the delay in the reception of the echoes of the transmitted pulse) and along with information on the materials' reflective property (considering the intensity of the reflected signal). However, it is still limited by the high running costs. Hence, the complementarity of optical/SAR/LiDAR measures can lead to a more comprehensive description of a surveyed scene if considering these data jointly. The differences among these three modalities can be seen at a glance by looking at Figure 1, in which a composition of the three acquisitions is presented.

The importance of fusing different modalities was already pointed out in many early works [1], [2] such as for the recognition of man-made objects by fusing LiDAR data and thermal images [3] or for scene interpretation [4] and image classification [5] when jointly considering optical and SAR images. Since the advent of remote sensing satellites, data fusion has been a very active field of research due to the increasing amounts of data available generated by the periodic acquisitions. So far, data fusion practices are currently widely employed in many applicative remote sensing tasks such as urban mapping [6], forest-related studies [7], [8], [9], oil slick detection and characterization [10], [11], disaster management [1], [12], and digital surface model (DSM) and digital elevation model (DEM) generation [13], to cite a few. Due to the ever increasing number of sensors operating with different characteristics and acquisition modalities, the potentialities and outcomes of data fusion are increasing. As a result, the interest of the remote sensing community around this topic keeps increasing. See for example the presence of active groups in professional societies dedicated to this topic (such as the IEEE Data Fusion Technical Committee and the Working

M. Dalla Mura is with the Univ. Grenoble Alpes, GIPSA-Lab, Grenoble, France

S. Prasad is with the University of Houston, USA

F. Pacifici is with DigitalGlobe Inc., CO, USA

P. Gamba is with the University of Pavia, Italy

J. Chanussot is with the Univ. Grenoble Alpes, GIPSA-Lab, Grenoble, France and with the Faculty of Electrical and Computer Engineering, University of Iceland, Iceland

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, Iceland

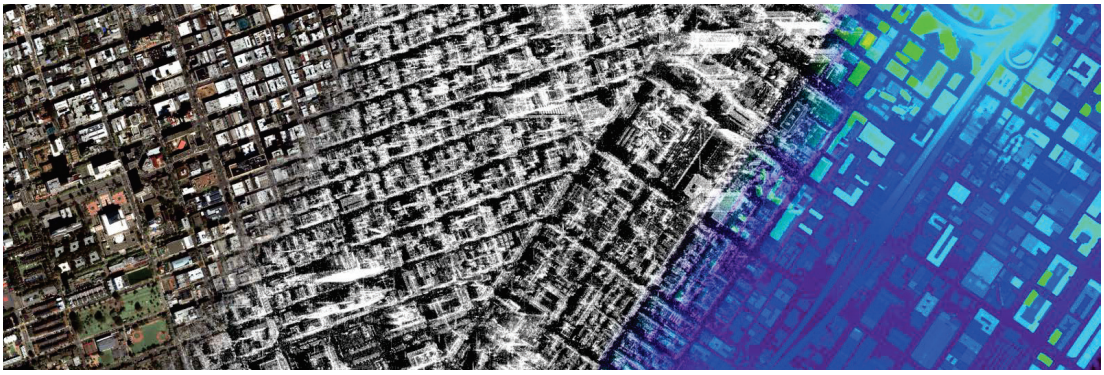


Fig. 1. From left to right, composition of an optical (true color composition with sub-meter spatial resolution [3 bands image]), SAR (amplitude of backscattering [scalar image]) and LiDAR elevation data [scalar image obtained by rasterizing the 3d point cloud] acquired over the city of San Francisco, U.S.A.. This set of data was used in the Contest of 2012. Source [16].

Group VII/6: Remote Sensing Data Fusion of the International Society of Photogrammetry and Remote Sensing), the constant presence of special sessions devoted to DF in almost all remote sensing conferences and workshops, or even entire conferences devoted to DF (such as the *International Symposium Remote Sensing and Data Fusion over Urban Areas*), and of special issues in remote sensing journals (e.g., the "Special issue on data fusion" of the *IEEE Transaction and Geoscience Remote Sensing* in 2008 [14] and the upcoming one of the *IEEE Geoscience and Remote Sensing Magazine* [15]).

Data fusion is a common paradigm related to the processing of data observed by different sensors and finds its place in a large variety of domains. Since a survey of the problem of DF in general terms is outside the scope of this paper, for reference we refer the interested reader to [17], [18], [19], [20]. If we focus on remote sensing, the approaches to data fusion are usually divided into three groups according to the level of the processing chain in which the fusion takes place [21], [22]. In general fusion can be performed at three different processing levels:

- **Raw data level** (also denoted as *Pixel level*). In some scenarios, the fusion of different modalities is performed at the level in which the data are acquired. The aim is in this case to combine the different sources in order to synthetize a new modality, which, afterwards, could be used for different applications. Image sharpening, super resolution and 3D model reconstruction from 2D views are examples of applications that share this aim.
- **Feature level**. The objective of DF at the feature level refers to the generation of an augmented set of observations considering data belonging to different sources. The result of the fusion can be taken jointly as input to a subsequent decision step. Focusing on land cover classification, perhaps the most straightforward way to perform this fusion is to stack one type of data on the other and to feed the classifier with this new data set. In other cases, different sets of features (e.g., image primitives such as linear features [23] or spatial features [24]) can be extracted from one or multiple data sources and combined together in order to reduce the uncertainty or achieve a richer description, respectively.
- **Decision level**. In this third case, the combination of the information coming from the different sources is performed on the results obtained considering each modality separately. If the data provide complementary information for the application considered, one can expect to increase the robustness of the decision through the fusion of the results obtained from each modality independently. This is achieved because in the result of the fusion the single decisions that are in agreement are confirmed due to their consensus, whereas the decisions that are in discordance are combined (e.g., via majority voting) in the attempt of decreasing the errors. The same concept can be found implemented by ensemble learning in pattern recognition [25].

This paper aims to present the current trends, opportunities and challenges of multimodal data fusion in remote sensing in the light of the outcomes of the IEEE Data Fusion Contests (DFCs) which have been taking place yearly since 2006.

The paper is organized as follows. A brief introduction of the nine contests issued from 2006 to 2014 are presented in Section II. Section III is devoted to present the applicative tasks of remote sensing in which data fusion approaches can be employed. Section IV proposes a discussion of the opportunities and challenges of data fusion in remote sensing and Section VI concludes this paper.

II. IEEE Data Fusion Contests

In order to foster the research on the important topic of data fusion, the Data Fusion Technical Committee (DFTC)[1] of the IEEE Geoscience and Remote Sensing Society (GRSS) has been annually proposing a Data Fusion Contest since 2006. The DFTC serves as a global, multi-disciplinary, network for geospatial data fusion, with the aim of connecting people and resources, educating students and professionals, and promoting the best practices in data fusion applications. The contests have been issued with the aim of evaluating existing methodologies at the research or operational level, in order to solve remote sensing problems using multisensoral data. The contests have provided a benchmark to the researchers interested in a class

[1] http://www.grss-ieee.org/community/technical-committees/data-fusion/

of data fusion problems, starting with a contest and then allowing the data and results to be used as reference for the widest community, inside and outside the DFTC. Each contest addressed different aspects of data fusion within the context of remote sensing applications. The contests proposed so far are briefly introduced in the following.

The focus of the 2006 Contest was on the fusion of images with different spatial and spectral characteristics [26] (see Sec. III-A for details on this application). Six simulated Pleiades images were provided by the French National Space Agency (CNES). Each data set included a very high spatial resolution monochromatic image (0.80 m resolution) and its corresponding multi-spectral image (3.2 m resolution). A high spatial resolution multi-spectral image was available as ground reference, which was used by the organizing committee for evaluation but not distributed to the participants.

In 2007, the Contest theme was urban mapping using SAR and optical data, and 9 ERS amplitude data sets and 2 Landsat multi-spectral images were made available [27]. The task was to obtain a classification map as accurate as possible with respect to the unknown (to the participants) ground reference, depicting land cover and land use patterns for the urban area under study.

The 2008 Contest was dedicated to the classification of very high spatial resolution (1.3 m) hyperspectral imagery [28]. The task was again to obtain a classification map as accurate as possible with respect to the unreleased ground reference. The data set was collected by the Reflective Optics System Imaging Spectrometer (ROSIS-03) optical sensor with 115 bands covering the 0.43-0.86 μm spectral range. Each set of results was tested and ranked a first time using the Kappa coefficient. The best five results were used to perform decision fusion with majority voting. Then, re-ranking was carried out after evaluating the level of improvement with respect to the fusion results.

In 2009-2010, the aim of the Contest was to perform change detection using multi-temporal and multi-modal data [29]. Two pairs of data sets were available over Gloucester, UK, before and after a flood event. The data set contained SPOT and ERS images (before and after the disaster). The optical and SAR images were provided by CNES. Similar to previous years' Contests, the ground truth used to assess the results was not provided to the participants.

A set of WorldView-2 multi-angular images was provided by DigitalGlobe for the 2011 Contest [30], [31]. This unique set was composed of five Ortho Ready Standard multi-angular acquisitions, including both 16 bit panchromatic and multi-spectral 8-band images. The data were collected over Rio de Janeiro (Brazil) in January 2010 within a three minute time frame with satellite elevation angles of 44.7°, 56.0°, and 81.4°in the forward direction, and 59.8°and 44.6°in the backward direction. Since there were a large variety of possible applications, each participant was allowed to decide a research topic to work on, exploring the most creative use of optical multi-angular information. At the end of the Contest, each participant was required to submit a paper describing in detail the problem addressed, the method used, and the final result generated.

The 2012 Contest was designed to investigate the potential of multi-modal/multi-temporal fusion of very high spatial resolution imagery in various remote sensing applications [16]. Three different types of data sets (optical, SAR, and LiDAR) over downtown San Francisco were made available by DigitalGlobe, Astrium Services, and the United States Geological Survey (USGS). The image scenes covered a number of large buildings, skyscrapers, commercial and industrial structures, a mixture of community parks and private housing, and highways and bridges. Following the success of the multi-angular Data Fusion Contest in 2011, each participant was again required to submit a paper describing in detail the problem addressed, method used, and final results generated for review.

The 2013 Contest aimed at investigating the synergistic use of hyperspectral and LiDAR data (in the form of LiDAR-derived digital surface model) that were acquired by the NSF-funded Center for Airborne Laser Mapping over the University of Houston campus and its neighboring area in the summer of 2012 [32], [33]. The 2013 Contest consisted of two parallel competitions: i) the best classification challenge and ii) the best paper challenge. The former was issued to promote innovation in classification algorithms, and to provide objective and fair performance comparisons among state-of-the-art algorithms. For this task, users were asked to submit a classification map of the data using the training samples generated by the DFTC via photo-interpretation. The validation set was kept unknown to the participants and used for the quantitative evaluation. The best paper challenge had the objective of promoting novel synergistic use of hyperspectral and LiDAR data. The deliverable was a 4-page manuscript that addressed the problem, methodology, and results. Participants were encouraged to consider various open problems on multi-sensor data fusion, and to use the provided data set to demonstrate novel and effective approaches to solve these problems.

The 2014 edition of the Data Fusion Contest proposed the fusion between images acquired at different spectral ranges and spatial resolutions [34]. Specifically, the data at disposal were a coarser-resolution long-wave infrared (LWIR) hyperspectral image (84-channels covering the wavelengths in the thermal domain between 7.8 and 11.5 nm with a 1m of spatial resolution) and a high spatial resolution data acquired in the visible (VIS) spectrum (RGB channels with a 20-cm spatial resolution) acquired over the same area. As for the Contest in 2013, two different challenges were proposed. One related to land cover classification and the other to a best paper challenge (i.e., leaving the application open).

III. Data fusion problems in remote sensing

This section aims at presenting the tasks pertaining to remote sensing treated by the Contests in which data fusion is employed.

A. Pansharpening

The so called Very High Resolution (VHR) satellites such as IKONOS, QuickBird and the more recent WorldView-2

and WorldView-3 are able to image a scene with *panchromatic* (PAN) and *multispectral* (MS) bands. The former is a monochromatic sensor acquiring the radiance of the scene in the Visible and Near InfraRed (VNIR) spectrum (typically in the interval 450-800 nm) with a sub-meter spatial resolution. The spatial resolution is measured in terms of Ground Sampling Interval (GSI) which is the distance on the ground between the centers of two adjacent pixels [35] and informally can be associated to the "pixel's size". Currently, the highest spatial resolution for commercial satellites is given by WorldView-3 with 0.31 m GSI at Nadir (i.e., direction perpendicular to the sensor) and 0.34 m at 20° Off-Nadir. The multispectral sensor acquires in different intervals of the electromagnetic spectrum thus providing an image composed of several spectral channels. The term spectral resolution is used in general for denoting the capability of the sensor in sensing the spectrum (number of spectral bands and width of the acquisition intervals in the spectral domain). The most typical configuration is four bands, (three in the visible, corresponding to the wavelengths of the red, green and blue colors and one in the near infrared domain) even if most recent sensors have expanded the number of channels. As an example, Figure 2 depicts the relative spectral responses of the sensors mounted on the Worldview-2 satellite. For comparison,

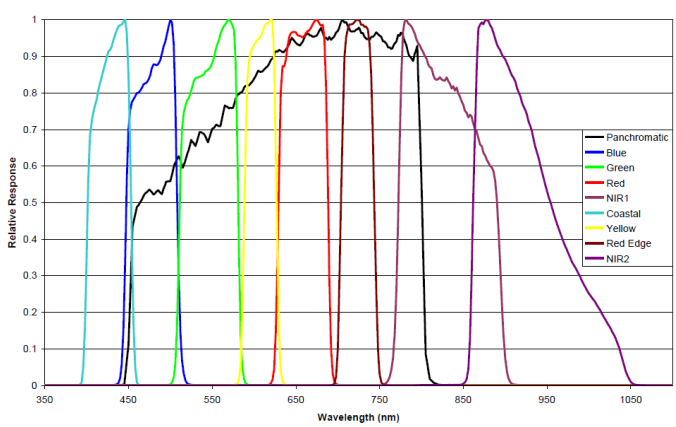


Fig. 2. Relative spectral responses of the sensors mounted on the Worldview-2 satellite.

the recent WorldView-3 acquires a 16-band product with 8 acquisitions in the VNIR and 8 in the Short Wave InfraRed (SWIR) spectrum. The GSI of the multispectral images is lower than the one of the panchromatic. This is due to a physical constraint that couples the spatial and spectral resolution and that prevents the arbitrarily reduction of the GSI simultaneously with the width of the spectral windows, (and the acquisition time) in order to guarantee a sufficient amount of energy reaching the sensor [35]. In general, the GSI of a multispectral band is a multiple of 4 with respect to the resolution of the panchromatic. For example for WorldView-3 the eight acquisitions in the VNIR spectrum have a GSI of 1.24 m at Nadir, 1.38 m at 20° Off-Nadir and in the SWIR Nadir of 3.72 m at Nadir and 4.10 m at 20° Off-Nadir.

Due to the above-mentioned physical limit in the acquisition, the PAN image shows a higher spatial resolution (i.e., a better capability in imaging the scene details) but a reduced spectral resolution (i.e., there is no chromatic information) with respect to the MS image. Since the common acquisition modality senses the scene both through the panchromatic and multispectral sensors simultaneously[2], a same scene is imaged in two products featuring complementary spatial and spectral resolutions. In the remote sensing community, the procedure aiming at synthesizing a new image with the spatial resolution of the panchromatic image, and the spectral resolution of the multispectral one is referred to as *Pansharpening* (i.e., the spatial sharpening of the multispectral channels through the use of the panchromatic image). This is clearly an instance of data fusion.

There is a constantly increasing demand for pansharpening products due to their use in many applications such as Earth visualization systems (e.g., Google Earth and Microsoft Virtual Earth) or as starting products in remote sensing applications such as change detection [36], object recognition [37] and visual image interpretation and classification [38]. Pansharpening presents some difficulties related to the fact that the details that are present in the panchromatic image appear blurred in the multispectral channels. Furthermore, such details would appear with variable intensity in the different spectral channels according to their spectral signature. This makes the retrieval of the single spectral contributions difficult due to the absent spectral information in the panchromatic image.

Many algorithms have been proposed in the literature of the last two decades, for detailed surveys the reader can refer to [39], [40], [41], [42]. The classical approach to pansharpening relies on the extraction of those spatial details from the panchromatic image that are not resolved in the multispectral one and their injection (appropriately modulated) into this latter one. This can be formulated as:

$$\widehat{\mathbf{MS}}_k = \widetilde{\mathbf{MS}}_k + g_k \mathbf{P}_D, \tag{1}$$

in which $\widehat{\mathbf{MS}}$, $\widetilde{\mathbf{MS}}$ and $\mathbf{P}_D$ are the result of pansharpening, with the MS image upscaled to meet the spatial resolution of the PAN and the spatial details of the PAN, respectively; k denotes the k-th spectral channel over N bands and $\mathbf{g} = [g_1, \ldots, g_k, \ldots, g_N]$ and g_k the injections gains. The way the operations of detail extraction and injection are performed determines the nature of the pansharpening algorithm. It is common practice to divide classical pansharpening algorithms into two families according to the technique used for estimating $\mathbf{P}_D$: the Component Substitution (CS) and the MultiResolution Analysis (MRA). The former extracts the details as:

$$\mathbf{P}_D = \mathbf{P} - \mathbf{I}_L \tag{2}$$

being $\mathbf{P}$ the PAN image and $\mathbf{I}_L$ a monochromatic image obtained by the weighted linear composition of the MS upsampled bands:

$$\mathbf{I}_L = \sum_{k=1}^{N} w_k \widetilde{\mathbf{MS}}_k. \tag{3}$$

[2]The delay between the two acquisitions can be considered negligible for typical remote sensing applications.

This approach can be equivalently implemented as a spectral transformation of the multispectral image into another feature space and on the subsequent substitution of one or more components in the transformed space with the PAN image followed by reverse transformation to produce the sharpened MS bands (hence the name CS). Some widely used algorithms based on this family are based on transformations such as Intensity-Hue-Saturation [43], [44], Principal Component Analysis and Gram-Schmidt orthogonalization [45].

The techniques belonging to the MRA class are based on the extraction of the spatial details present in the panchromatic image (and not fully resolved in the multispectral one) and their subsequent addition to the MS bands. Thus $\mathbf{P}_D$ here is computed as:

$$\mathbf{P}_D = \mathbf{P} - \mathbf{P}_L, \tag{4}$$

with $\mathbf{P}_L$ a low pass version of the PAN image obtained by spatially filtering $\mathbf{P}$. The spatial details can be extracted by several approaches as using an average filter [44], [35] or multiresolution decompositions of the image based on Laplacian [46] pyramids, or wavelet/contourlet operators [47], [48].

For both families, the injection of spatial details into the interpolated MS bands are weighted by gains (g_k) different for each band and either considering them constant for each channel of varying locally (i.e., leading to "global" or "local" approaches, respectively). Pansharpening techniques based on the paradigm in Eq. 1 differ according to the way they compute $\mathbf{I}_L$ for CS techniques (i.e., how are the weights w_k in Eq. 3 obtained), $\mathbf{P}_L$ for MRA ones and the injection gains g_k.

The validation of the results in the context of pansharpening cannot be performed directly since there is no reference data. For this reason several attempts have been made for assessing quantitatively the results of pansharpening. Two validation strategies are mostly used. The first is based on the reduction of the spatial resolution of both the original MS and PAN images and then the original MS image is used as reference for the evaluation of the results [26]. The underlying assumption in this strategy is that the tested algorithms are invariant among resolutions [49]. However, this hypothesis is not always verified in the practice, especially for very high resolution images acquired on urban areas [50]. The full scale validation employs indexes that do not require the availability of a reference image since they evaluate the relationships, such as the spectral coherence, among the original images and the pansharpened product [50], [51]. In this case the evaluation is done at the native scale of the images but clearly the results depend on the definition of such indexes.

We leverage the results of the DF Contest issued in 2006 [26] for bringing about a discussion on the performances of different pansharpening algorithms. In this contest, the participants were asked to perform pansharpening on a set of simulated images from the Pleiades sensor and a spatially downsampled image acquired by QuickBird. Each data set included VHR panchromatic image and its corresponding multispectral image. A high spatial resolution multispectral image was available as ground reference, which was used by the organizing committee for evaluation but not distributed to the participants. This reference image was simulated in the Pleiades data set and it was the original multispectral image in the QuickBird one. The results of the algorithms submitted by the different research groups were compared with a standardized evaluation procedure, including both visual and quantitative analysis. The former aimed at comparing the results in terms of general appearance of the images as well as by means of a local analysis focusing on the rendering of objects of interest such as linear features, punctual objects, surfaces, edges of buildings, roads, or bridges. The quantitative evaluation was performed using quality indexes for measuring the similarity of the fused results with respect to the reference image.

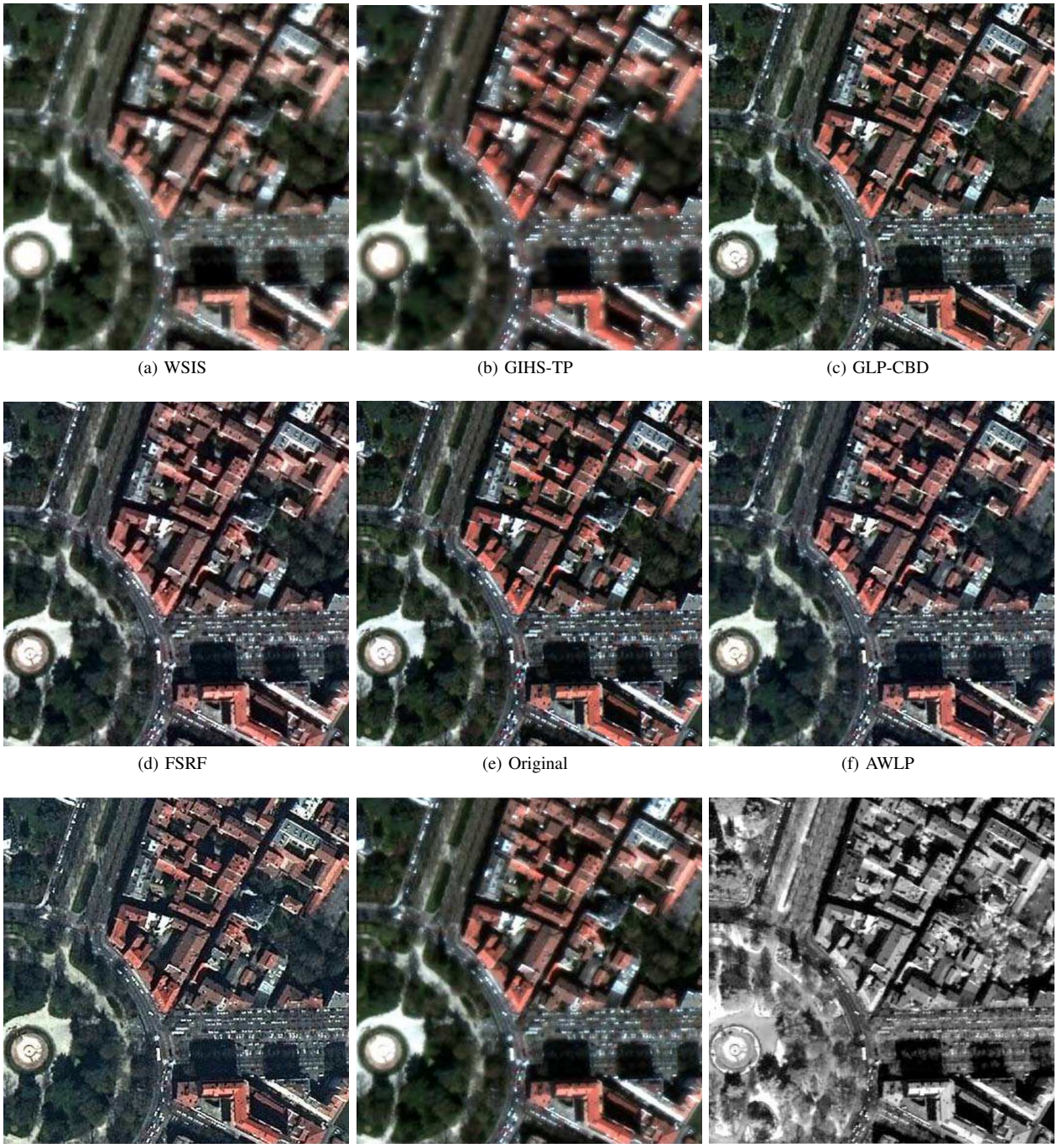
Examples of pansharpening results submitted to the contest are shown in Figure 3.

As it is possible to notice by looking at the figure, the products of the fusion present differences in terms of both radiometry (e.g., color) and geometry (i.e., rendering of the spatial details). Relying on their evaluation (reported in [26]), it is possible to draw some concluding remarks. CS techniques yield in general fused products with accurate spatial details since no spatial filtering is performed (the low resolution PAN is estimated from the MS image according to Eqs. 2,3), but can often produce spectral distortions which can be seen in the fused images as a too high or low saturation of a certain color component. The results obtained by MRA methods typically better preserve the spectral content, but at the detriment of the spatial fidelity of the details. Indeed the spatial filtering for extracting the details to inject can in some cases produce spatial artifacts or blurred areas according to [50]. Among the algorithms considered in the contest, the best results (both in terms of visual and quantitative analysis) were obtained by two algorithms from the MRA family: GLP-CBD and AWLP in Figure 3. These two pansharpening techniques extract the spatial details with a multiresolution decomposition of the PAN (Eq. 4) with a Gaussian pyramid for the former and wavelet filters for the latter. It is worth emphasizing that even if the two filters are different, their frequency response is very similar and it can be seen as an approximation of the Modulation Transfer Function of the sensor (i.e., the transfer function of the optical system [35]). This is a fundamental aspect since selecting a filter that models as closely as possible the blur that relates the MS and the PAN sensor, it is possible to obtain an accurate extraction of the spatial details and consequent consistent pansharpening result.

For a more comprehensive comparison among several pansharpening algorithms the reader is referred to [42].

B. Change detection

Change Detection (CD) refers to the task of analyzing two or more images acquired over the same area at different times (i.e., multitemporal images) in order to detect zones in which the land cover type changed between the acquisitions [52], [53], [54], [55], [56]. There is a wide range of applications in which change detection methods can be used, such as urban and environmental monitoring, agricultural and forest surveys, and disaster management. In general CD techniques



(a) WSIS (b) GIHS-TP (c) GLP-CBD

(d) FSRF (e) Original (f) AWLP

(g) UNP-PanSharp (h) GIHS-GA (i) Panchromatic

Fig. 3. Results of the 2006 Data Fusion contest on pansharpening (pansharpening family reported in parenthesis): (a) Weighted Sum Image Sharpening, WSIS (CS); (b) Generalized Intensity Hue Saturation With Tradeoff Parameter, GIHS-TP (CS); (c) Generalized Laplacian Pyramid With Context-Based Decision, GLP-CBD (MRA); (d) Fast Spectral Response Function, FSRF (CS); (e) Original image used as reference in the validation; (f) Additive Wavelet Luminance Proportional, AWLP (MRA); (g) University of New Brunswick (UNB)-Pansharp (CS); (h) Generalized Intensity Hue Saturation With Genetic Algorithm, GIHS-GA (CS); (i) Panchromatic image. Source [26].

assume multitemporal images to be captured from the same sensor and possibly with same acquisition modality (e.g., angle of view) in order to reduce the problems of co-registration between images and minimize the presence of differences in the images that are not due to a real change in land cover. In the case of natural disasters and search and rescue operations, where time is a constraint and the data available is usually fragmented, not complete, or not exhaustive the analysis has to be performed using images acquired from different sensors. Thus, CD encounters greater challenges and its accuracy relies on the way the different modalities are handled.

In the following, we will briefly introduce the main approaches that have appeared in the literature for performing CD and we will focus on CD based on different modalities. CD

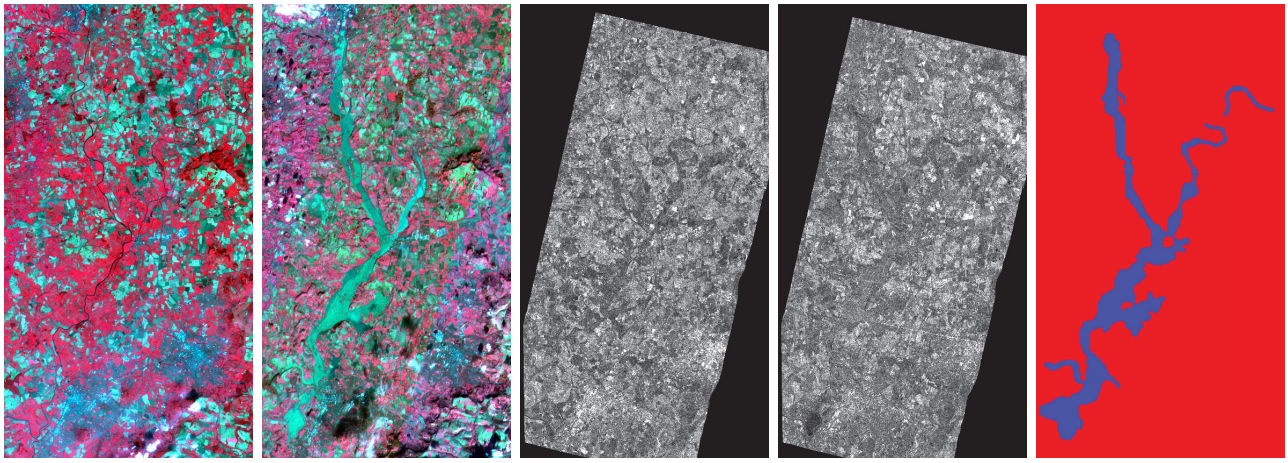
can be seen as a particular instance of thematic classification of the land cover, in which the classes are change and no-change.

The methods proposed in the literature can be divided into two main approaches: i) supervised and ii) unsupervised CD. The first relies on the presence of a priori information on the scene such as examples of changed and unchanged areas. This information could be derived from field surveys or defined by the user through photo-interpretation. The availability of labeled information allows one to perform the detection of land cover transition employing conventional supervised classification techniques. Two main approaches are presented in the literature according to the stage of the CD process in which the classification step is performed: *post-classification comparison* [53], in which classification is done independently at each acquisition, and the changes are then detected from a comparison of classification maps; and *multi-date classification* [52], where multi-temporal information is considered simultaneously for classification. Semi-supervised approaches also exist and have recently gained interest from the community since they handle the lack of labeled information for some dates, which might be a frequent operational scenario. These techniques are in general based on transfer learning and domain adaptation methods (such as [57]). The advantage of supervised technique lies in the fact that the analysis is built on the definition of change. Moreover, if the labeled information comprises information on different land cover types, the analysis can also determine the type of change according to the type of land cover transition that occurred. However, these approaches also have some drawbacks due to the classification step, for example CD results can be affected by misclassification errors (especially for techniques based on post-classification comparison) [29]. In addition, these techniques are limited by the availability of labeled samples. Unsupervised approaches to CD do not require any ground reference and will detect changes as (in general sudden) variations in the evolution of land covers. In general, these techniques detect only the presence of changes [56]. Recently, in specific cases some techniques have been proposed for discrimination among different types of changes [58], [59]. However, the detected change cannot be associated with thematic information (e.g., on the type of land-cover transition) since no reference on the ground is available. Unsupervised techniques attempt to detect variations in land covers based on some dissimilarity measures (e.g., multivariate differences [56]) computed among the images acquired at different dates or statistical tests (e.g., [60]). With a focus on CD performed on optical images, the change is related to a variation in the radiometry of the scene, which refers to the values of radiance captured by the sensor. Changes of interest are usually related to variations in radiance that are related to a change in the reflectance of the land cover rather than to variations due to differences in the acquisition settings such as illumination changes, different data normalization and calibration settings [56]. In order to cope with these latter sources of radiometric variations and detect the relevant changes, the multivariate alteration detection (MAD) technique with iterative reweighted (IR-MAD) scheme [61], [62] was proposed. When considering data acquired by different modalities, capabilities in providing a fast response can greatly improve. However, using different data belonging to sources that might be significantly different can be a severe issue to handle. Comparison between modalities can be meaningless if not done appropriately, differences in acquisitions can become prohibitive for the generation of consistent results. In 2009-2010, the contest was issued to address the task of CD using multi-temporal and multi-modal data [29]. See Figure 4 for the dataset used in the contest.

The two pairs of data sets made available to the participants were acquired before and after a flood event. The class "change" was the area flooded by the river and the class "no change" was the ground that had not been concerned by the flooding. The optical and SAR images were provided by CNES. The participants were allowed to use supervised or an unsupervised method with all the data, the optical data only, or the SAR data only. A variety of supervised and unsupervised approaches were proposed by the participants. Interestingly, a simple unsupervised change detection method resulted in similar classification accuracies compared with supervised approaches. As expected, the approaches that utilized both SAR and optical data outperformed other approaches, although the contribution of SAR data alone was minimal to the overall change detection accuracy (due to the high discrimination capability of the optical data for this task). The overall best results were obtained by fusing the five best individual results via majority voting. Remarkably, considering both SAR and optical data jointly in an unsupervised scheme led to slightly degraded performances with respect to the use of only optical data. In regard to this result, we remark that the analysis was performed with an unsupervised approach, preventing the analysis to target closely the objective of the task as for a supervised approach, in which the available a priori information is exploited.

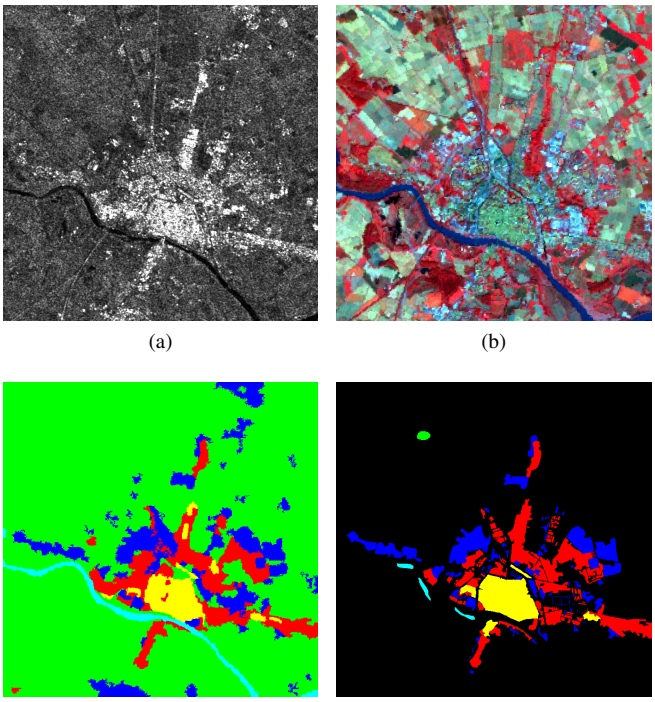
C. Classification

Various past contests have focused on the fusion of data in order to provide superior classification accuracy (compared to considering the single modalities only) for remote sensing applications. Previous contests have provided other multi-modality fusion scenarios — both in terms of sensors and challenges (e.g. use of optical imagery, LiDAR data, SAR data etc.) for various image classification scenarios [27], [28]. We take the most recent one — the 2013 contest involving multi-sensor (hyperspectral and LiDAR) for urban classification, as an example to highlight emerging trends. This contest saw a very wide range of submissions — utilizing hyperspectral only, or using hyperspectral fused with LiDAR in the original measurement domain or in feature spaces resulting from spatial and other related features extracted from the dataset. Submissions that provided high classification performance often utilized LiDAR data in conjunction with the hyperspectral image, particularly to alleviate confusions in areas where the spectral information was not well-posed to provide a good solution (e.g., classes that had similar material compositions but different elevation profiles), and vice-versa.



(a) (b) (c)

Fig. 4. Dataset of the 2009-2010 Contest. Color composition of the SPOT optical image (a) and ERS single amplitude SAR data (b), collected before (left) and after (right) the flood event, provided as input to the change detection problem. The reference map used for the evaluation of the submitted algorithms is shown in (c). Source [29].



(a) (b) (c) (d)

Fig. 5. Dataset of the 2007 Contest. City of Pavia imaged by (a) SAR (backscattering amplitude) and (b) optical (bands RGB-431) sensors. In (c) and (d), the final classification map and the ground reference data are shown. Source [27].

Another focus area of emerging and promising contributions to the range of submissions, involved post-processing of classification results to mitigate salt-and-pepper errors in classification. We note that this classification contest was designed to pose some unique challenges — specifically, the training mask and test masks were spatially disjointed, and had substantial variability. Some classes existed under a cloud shadow in the testing masks, testing algorithms while other were submitted for their capability to adapt to such variations. Most submissions did not fare well under cloud shadows, but submissions where contestants utilized spatial contextual information fared much better in general, even under cloud shadows. The winning algorithm was based on spectral unmixing, and utilized abundance maps derived from hyperspectral imagery as features, in conjunction with raw hyperspectral and LiDAR data, using Markov Random Fields and ensemble classification. As a general trend, we have seen a great degree of variability between classification performance of various methods submitted for data fusion and classification — be they feature level fusion or decision level fusion. It is difficult to identify any one method that performs well in general — to a great degree, this depends on the underlying problem and the nature of the datasets.

With that background, we next summarize some emerging trends in the general area of classification for multi-modality data fusion for remote sensing. We recognize that as in many application domains, "classification" implementations take the following flow — (1) Preprocessing and feature extraction followed by classification. Pre-processing steps refer to operations undertaken to better condition the data prior to analysis. These include spectral-reflectance estimation from at-sensor radiance for hyperspectral measurements (e.g. using atmospheric compensation techniques that rely on physics based models [63] or statistical models [64]); geo-registration of multiple modalities, spectral radiance/reflectance denoising etc. Reflectance estimation is crucial when utilizing prior libraries that have been constructed *outside* of the current scene being analyzed, accurate geo-registration is critical in multi-modality frameworks, denoising is helpful when utilizing spectral imagery at longer wavelengths etc. Feature extraction is often a critical preprocessing technique for the classification of single and multi-modality image analysis. With modern imagers (e.g. hyperspectral), the resulting dimensionality of feature spaces is intractably high. This has ramifications wherein classification algorithms struggle to estimate statistics (or overfit) when using raw data. A variety

of linear and nonlinear feature extraction algorithms exist to alleviate this problem, with the end goal of transforming this data to a lower dimensional subspace better conditioned for classification. These can be categorized into feature selection approaches [65], [66], feature projection approaches [67], linear and nonlinear approaches, and supervised, unsupervised, or semi-supervised approaches. An emerging area within the feature extraction category is *nonlinear manifold learning* that recognizes that high dimensional remote sensing data often resides in a lower dimensional manifold — techniques that characterize and learn the manifold structure from training data have been shown to yield superior features for classification, pixel unmixing and data fusion tasks [68].

While nonlinear support vector machine classifiers and their many variants have gained popularity in the remote sensing community, a variety of classification approaches are now prevalent. Among these include approaches that rely on statistical models [69], sparse representation models [70] etc. We note that among these methods, statistical classifiers (e.g. the Gaussian mixture model) are extremely sensitive to the dimensionality of the data, and hence a feature reduction scheme is often employed as a preprocessing technique for such classifiers. Within the realm of supervised classification for remote sensing, active learning is a potentially useful paradigm — with ground data being expensive (and in many cases difficult) to acquire, a strategic sampling scheme is desirable. Active learning provides a closed-loop (annotator-in-the-loop) framework whereby the classifier guides collection of strategic field samples that add the most value to the underlying classification task. These approaches have been developed and optimized for various classifiers for remote sensing image analysis [71].

We note that several of these approaches have been recently extended to multi-modality or multi-source image analysis frameworks. For instance, in [72] a composite kernel SVM was implemented for multi-source data fusion; in [73] a composite kernel local Fisher's discriminant analysis was implemented (CK-LFDA) for multi-source feature extraction in a kernel induced space wherein a composite kernel feature space was constructed that optimally represented (in the sense of the local Fisher's ratio) multi-source data; [74] provides a framework for multi-source active learning using multi-kernel learning, etc. Likewise, statistical classifiers have been used for effective data fusion for remote sensing image analysis [75].

The emerging paradigms of deep learning provide an approach to systematically and hierarchically learn the underlying structure in datasets via deep neural networks [76], [77]. In recent years, deep hierarchical neural models have been proposed to learn a feature hierarchy — from input images to the back-end classifier. Typically in such architectures, image patches are convolved with filters, responses repeatedly subsampled, and refiltered — when passed through sufficient layers of convolution, subsampling (and nonlinear mapping through activation functions), it is expected and observed with real data that the resulting feed forward network is very effective for image analysis. Although deep learning has been successfully applied to many computer vision applications, its utility for single and multi-sensor remote sensing data has been very limited, although the potential benefits to multi-sensor data fusion are enormous.

D. Miscellaneous applications

As mentioned in Sec. II, the most recent Contests accepted submissions in which the objective of the fusion was not imposed in order to encourage new applications. This was done for exploring the capabilities in using the data provided in the framework of the contests in unforeseen problems. Besides the "regular" data fusion tasks discussed previously, a number of interesting research topics were proposed and addressed demonstrating numerous possibilities and a variety of applications that multi-modal remote sensing images can offer.

For instance, hyperspectral and LiDAR data, and depth images at different locations are used in [78] to quantify physical features, such as land-cover properties and openness, to learn a human perception model that predicts the landscape visual quality at any viewpoint. Techniques to track moving objects (such as vehicles) in WorldView-2 images are illustrated in [79] and [80]. The main idea is based on the time gap between different banks of filters. Radiosity methods are discussed in [16] to improve surface reflectance retrievals in complex illumination environments such as urban areas, whereas [33] presents a methodology for the fusion of spectral, spatial, and elevation information by a graph-based approach. Other contributions include methods to derive an urban surface material map to parameterize a 3-dimensional numerical microclimate model, to retrieve building height [81], to applications such as visual quality assessment and modeling of thermal characteristics in urban environments. Likewise, another proposed work was a new method that focused on removing artifacts due to cloud shadows that were affecting a small part of the image [32].

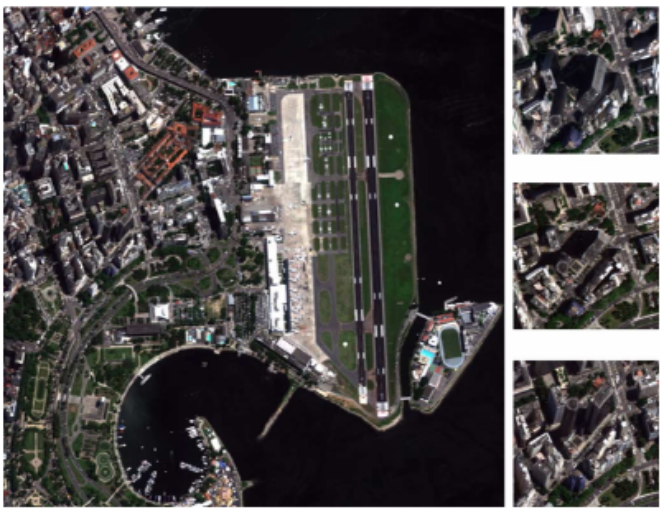


Fig. 6. The WorldView-2 scene provided for the 2011 Data Fusion Contest with three details from the three most nadir-pointing images. Source [27].

IV. Discussion

In this section we want to highlight some relevant aspects of data fusion in remote sensing by leveraging the outcomes

of the contests. As introduced in Sec. I and seen as in practice from the challenges proposed, data fusion can take place at different levels in the generic scheme aiming at extracting information from data.

- **Raw data level**. Examples of applications considered in the contests in which fusion was performed at this level are pansharpening (Sec. III-A) and DSM generation from multiangular images (e.g., Figure 6). Usually in these specific tasks there are some constraints that bound the analysis. Particularly, it is possible to rely on some similitudes among the data to fuse. For example when considering the analysis of multiangular data the sensor used in the acquisitions is the same. In the specific case of the scenario of the contest of 2011, the images were acquired in a single pass of the satellite; hence limiting the variations in the images due to different illumination condition (as it would be the case for acquisitions done at different dates). Analogously, for pansharpening, the panchromatic and multispectral sensors are mounted on the same platform (this makes the spatial registration between images not necessary) and with a negligible time lag. This applies also to other tasks that were not presented in this paper such as in hyperspectral imaging for combining spectral channels (for generating a new image with a different configuration of the spectra), spectral and spatial features (hyperspectral images) [82].
- **Feature level**. Fusion at the feature level took place in several proposed techniques addressing tasks such as classification and change detection. Features were extracted by one or more modalities and subsequently fused in order to compose a new enriched set of characteristics. Demonstrations of fusion on a single modality are are given for example, when combining spectral with spatial features. In this case, in order to properly perform such fusion, the differences between the modes should be taken into account in order to be able to properly exploit them. For example, in the context of classification with LiDAR and optical images, if one wants to use both sources as input to a classifier, then registration problems should be solved (e.g., by rasterizing the LiDAR data to the same spatial resolution of the optical image).
- **Decision level**. Fusion of decisions occurs at the highest semantic level. Among the contests, we recall that the one of 2008 [28] was based on such DF paradigm (i.e., ranking the submitted classification maps on the basis of their amount of relative contribution in the final decision obtained by majority voting on them). Decision fusion took place also in other contests both performed by the contests' organizers (such as in 2009-10 [29]) as by in some of the techniques proposed by the participants. According to the results, DF at this level proved to be very effective even with a simple fusion strategy such as the majority voting.

By looking at the results of this review it is possible to make some general remarks:

- For certain applications, the exploitation of multiple modalities through a DF paradigm is the sole way for performing the analysis. This is the case when the fusion takes place at the raw level. For example, it would not be possible to derive a 3D model of any scene only with a single acquisition. Moreover, it is only through the joint consideration of multisensoral data that it is possible to observe some phenomena (e.g., for the retrieval of biophysical parameters which cannot be sensed by using the acquisitions of a single sensor or single modality [83]). Likewise, this more complete description of the observed world can make certain operations possible. In classification the discrimination between several classes might only be possible if multimodal data is considered. For instance, LiDAR gives information on the elevation of the objects in a scene, while a multispectral sensor captures the spectral properties of the materials on their surfaces. Clearly, land cover types differing in both of these characteristics could not be discriminated by considering only one of these modalities.
- It is necessary to consider the sensors and data characteristics. Especially when the data show extremely different resolutions or significantly different geometries in the acquisition. For example, by considering a fusion between a SAR and an optical image, the position in a SAR image of the contributions of the objects in a scene is dependent on their distance to the sensor, whereas an optical image reflects their position on the ground. In addition, the SAR image can show patterns (such as those due to double bounce, layover and shadowing effects) that find no correspondents in the optical image. In this case, a trivial pixelwise combination of a VHR optical and SAR image might lead to meaningless results. The joint exploitation of the two modalities can only take place if one properly accounts for the model describing the way the acquisitions are done and if a 3D model of the scene is available [16]. Analogously, the more knowledge of the sensors is included in the analysis, the better the accuracy of the fusion results. As shown for pansharpening (Sec. III-A), the more precise and meaningful results where obtained by taking into consideration the blur that models the difference in terms of spatial resolution between the panchromatic and the multispectral acquisition. In addition, DF should considered *cum grano salis* since the data characteristics are not properly accounted for if the a priori information (e.g., given by the application) is not included. Related to this latter aspect, we remark how fusing different data can even prevent the correctness of the results (e.g., as reported in Sec. III-B for change detection in a completely unsupervised mode). Thus, considering data that are not relevant for the application could even harm the analysis. So this last aspect opens some questions on the motivation of the fusion, since considering a fusion of different modes further increases the complexity of the system and the computational burden. So the use of different modes should be supported by its actual need. In order to address this last aspect, a priori information on the application and a knowledge of the characteristics of the different modalities should be considered in advance.

- Despite the clear benefits that data fusion can bring, it can lead to some important challenges. Data acquired from different sources might come in completely different formats. For example, imaging sensors provide data over a lattice, whereas LiDAR generates a set of sparse and non-uniformly spaced acquisitions. In addition, pixels in optical images and data in LiDAR are multivariate real values whilst radar images have complex values. Having to convert the data into common formats for processing them jointly can generate additional uncertainty in the measure (e.g., greater errors due to operations such as quantification and interpolation) in the data with respect to the one inherent to each single modality. When the correspondence among multisensoral data cannot be established, the result of the fusion might present missing information (for some modalities). This creates theoretical and algorithmic challenges related to the way missing data are handled.

V. Perspectives

From the current status resumed in the previous section, here we will account for still open challenges and new perspectives of DF in remote sensing.

- The number of new satellites that are planned to be launched in the near future is constantly increasing, and companies such as Planet Labs and SkyBox are building Earth observation constellations of hundreds of satellites. In addition to this increasing trend, satellite platforms are getting more and more diversified in terms of characteristics. For example, the recently launched WorldView-3 by DigitalGlobe includes 29 bands in the VNIR-SWIR region of the spectra, ranging from 30 cm to 30 m resolution. These two aspects are leading to the generation of data acquired by a plethora of different sensors, that will consequently produce an increasing need of DF analysis in order to fully exploit such data. In this perspective, we can think of DF approaches that are less sensor dependent have an advantage, due to a larger application scope.
- Another current trend we are witnessing is the improvement of the sensors' resolution (geometric, spectral or radiometric). This is surely a very favorable feature, but it induces an increasing effort in the analysis [84]. Higher resolution data are able to sense more finely a scene (i.e., provide more geometric/spectral/radiometric details) increasing the amount of information, meaningful for a given task, that can be extracted but making the process of processing it from data more complex. This applies to each single modality, so when fusing multiple information sources the potentialities and difficulty in mining the relevant information scale accordingly. Furthermore, due to the increasing presence of satellite constellations providing larger coverage on the Earth's surface with smaller revisit times and the availability of archive data, a potentially massive amount of data could be processed. The need of efficient algorithms able to cope with such large amount of data will increasingly be a demand for new DF approaches that should be used in operational scenarios.
- In the last years, the remote sensing market has not only been exclusively considering data acquired by large satellites launched by governmental space organizations or large EO companies. Technological advancements have permitted to produce miniaturized satellite platforms [85] such as micro (10-100 kg), nano (1-10 kg), pico (0.1-1 kg) and even femto satellites (< 0.1 kg) [86]. Since smaller platforms have costs that are dramatically lower (about 10^3 [85]) with respect to those required by large ones (> 500 kg), launching satellite EO instruments has become an affordable business even accessible to universities [87]. Furthermore, terrestrial Unmanned Aerial Vehicles (UAVs) have largely spread, becoming undoubtedly an asset also for EO applications [88]. The low cost of off-the-shelf flying platforms and the possibility to equip them with consumer-level instruments (e.g., compact cameras) have made EO accessible to a larger amatorial public. Such increasing presence of small satellites and terrestrial EO platforms, is going to provide a consequently large amount of diversified data further broadening the scenario in which DF can take place. Clearly the instruments of the payload (e.g., sensors, GPS receivers and inertial systems) are significantly poorer in terms of performances with respect to those mounted on large satellites or professional airborne acquisition systems. The lower quality of these data can directly affect DF results if not properly accounted for in the analysis.
- In a larger perspective, we also envisage that additional information sources will be exploited for fusion in the near future. So far DF for EO has been almost exclusively based on remote sensing data (e.g., active and passive imagery). For example, the information available from Geographic Information Systems (GISs), such as road networks, building footprints from cadastral layers, land cover maps etc. can be of fundamental importance for EO applications. Some examples of fusion between remote sensing data and GIS layers have been made (e.g., [2]) but it has not taken off extensively in the remote sensing research community, even if, coupling the two information sources has been proven to be successful, being a standard technology in many EO visualization systems such as Google Earth. Perhaps a reason for the lack of established DF techniques for fusing GIS layers and remote sensing data can be attributed to the inherently different features of the two sources. In many cases, GIS layers can not be fused straightforwardly with remote sensing data, since they come in vectorial format, they might contain descriptive data (i.e., geo-localized textual information), they deal natively with semantic objects instead of pixels and coregistration with remote sensing images can be a severe issue. However, GIS data are largely available thanks to local information systems held by municipalities or by worldwide databases such as OpenStreetMap. In addition, some information issued from GIS can cover periods before the first acquisitions

available in remote sensing archives, hence becoming the only information available for some applications. Due to their large availability and features complementary with conventional remote sensing imagery, we believe there will be an increasing push in fusing GIS and remote sensing data.

In the same research direction, we also think that less conventional sensors could be beneficial for some applications. For example, the use of ground information (e.g., images from mobile phones, street views, height values from GPS sensors), opportunistic sensors (e.g., as passive sensors based on GPS signals [89]) or even geographically distributed sources that are not strictly remote sensing data, such as geo-tagged tweets, locations extracted from news, track points, etc. Considering these heterogeneous data together will definitely be a new challenge for DF.

- The ever increasing heterogeneity (in terms of resolution, characteristics, sources and consistency) of the data available for fusion will greatly influence the methodological development of DF algorithms, which in our opinion, will be increasingly application-driven. In fact, it will be unlikely that general purpose DF strategies will be able to deal with the different characteristics of data and be sufficiently valid for several tasks. The specific applications will define which sources are relevant for the fusion and how to combine them. As an example, in the framework of urban remote sensing, the model of the urban area (e.g., building, district or town level) or the phenomenon under study (e.g., detection of urban heat islands, or air pollution) will be constrained implicitly or explicitly which data to use and at which spatial and temporal scale.
- With a particular regard to the way to handle imperfect, missing and conflicting data, it is evident from the remarks in Sec. IV, that there is a lack of a universally recognized framework in which to perform the fusion properly taking into account these different characteristics [20]. This problem has been partially addressed by probability theory, fuzzy set theory, possibility theory, rough set theory, and Dempster-Shafer evidence theory but none of these approaches have been used extensively in DF problems [20], making this a still open challenge. This aspect will surely increase in importance since, as mentioned previously in this section, due to the emerging trends in EO there will large amounts of data available with heterogeneous characteristics and qualities.
- Another fundamental challenge is related to the validation of the results. This is an perennial problem for tasks in which there is no reference available (e.g., pansharpening) [82], [90]. However, the availability of commonly recognized validation paradigms is essential for evaluating newly proposed algorithms in a quantitative way.

VI. Conclusion

By reviewing the outcomes of the contests issued by the DFTC, it is possible to remark their main contributions such as i) fostering the methodological development on the topics defined by each contest; ii) making datasets available to the community — sometimes such datasets were valuable because it might be unusual in real operational scenarios to have so many data acquired from different sensors with such high spatial resolution, such as for the contests 2012 (VHR SAR, VHR optical from different sensors and LiDAR) and 2011 (images VHR multiangular) and iii) encouraging the emergence of new applications or research directions based the data of the contests. From the analysis of the different aspects of data fusion in remote sensing through the lens of the contests, one can clearly state that data fusion is indeed the way to extract information. Indeed, for some tasks and applications it is the sole mean to perform the analysis.

By taking as an example classification, which perhaps is one of the most active tasks that can be impacted by data fusion: with respect to one goal (partition of the data into a number of classes of interest), data fusion can lead to improved classification performances by providing complementary information, by reinforcing our belief in a result, or by solving ambiguities/conflicting situations. This is especially useful for the problem of land-cover/land-use mapping in a variety of applications. Multimodality can also be beneficial in a number of other situations, trying to provide a better description of the physical real-world. Each modality provides one projection of the complex physical world. Using multimodality is a way to access this complexity in a refined way, but combining these projections in an efficient and reliable way is a challenge. Specifically, as we saw in the preceding discussion, data fusion also presents several unique challenges both from the technical and methodological points of views, necessitating continued investigation from the research community.

Acknowledgment

This work is partially funded by the European project ERC-2012-AdG-320684-CHESS.

References

[1] T. Bell, "Remote sensing," *Spectrum, IEEE*, vol. 32, no. 3, pp. 24–31, Mar 1995.

[2] C. Pohl and J. Van Genderen, "Review article multisensor image fusion in remote sensing: concepts, methods and applications," *International Journal of Remote Sensing*, vol. 19, no. 5, pp. 823–854, 1998.

[3] C.-C. Chu and J. Aggarwal, "Image interpretation using multiple sensing modalities," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, no. 8, pp. 840–847, Aug 1992.

[4] V. Clement, G. Giraudon, S. Houzelle, and F. Sandakly, "Interpretation of remotely sensed images in a context of multisensor fusion using a multispecialist architecture," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 31, no. 4, pp. 779–791, Jul 1993.

[5] A. Solberg, A. Jain, and T. Taxt, "Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 32, no. 4, pp. 768–778, Jul 1994.

[6] P. Gamba, "Human settlements: A global challenge for EO data processing and interpretation," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 570–581, March 2013.

[7] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 5, pp. 1416–1427, 2008.

[8] S. Delalieux, P. Zarco-Tejada, L. Tits, M. Jimenez Bello, D. 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, June 2014.

[9] A. Ghulam, "Monitoring tropical forest degradation in Betampona nature reserve, Madagascar using multisource remote sensing data fusion," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. PP, no. 99, pp. 1–12, 2014.

[10] C. Brekke and A. H. Solberg, "Oil spill detection by satellite remote sensing," *Remote Sensing of Environment*, vol. 95, no. 1, pp. 1 – 13, 2005.

[11] M. Fingas and C. Brown, "Review of oil spill remote sensing," *Marine Pollution Bulletin*, vol. 83, no. 1, pp. 9 – 23, 2014.

[12] F. Dell'Acqua and P. Gamba, "Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives," *Proceedings of the IEEE*, vol. 100, no. 10, pp. 2876–2890, Oct 2012.

[13] F. Remondino and S. El-Hakim, "Image-based 3d modelling: A review," *The Photogrammetric Record*, vol. 21, no. 115, pp. 269–291, 2006.

[14] P. Gamba, J. Chanussot *et al.*, "Foreword to the special issue on data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1283–1288, 2008.

[15] "Special issue on "data fusion in remote sensing"," *Geoscience and Remote Sensing Magazine, IEEE*, vol. 2, no. 3, pp. 57–57, Sept 2014.

[16] C. Berger, M. Voltersen, R. Eckardt, J. Eberle, T. Heyer, N. Salepci, S. Hese, C. Schmullius, J. Tao, S. Auer, R. Bamler, K. Ewald, M. Gartley, J. Jacobson, A. Buswell, Q. Du, and F. Pacifici, "Multi-modal and multi-temporal data fusion: Outcome of the 2012 GRSS data fusion contest," *IEEE JSTARS*, vol. 6, no. 3, pp. 1324–1340, June 2013.

[17] D. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, Jan 1997.

[18] D. L. Hall and S. A. McMullen, *Mathematical techniques in multisensor data fusion*. Artech House, 2004.

[19] J. Esteban, A. Starr, R. Willetts, P. Hannah, and P. Bryanston-Cross, "A review of data fusion models and architectures: towards engineering guidelines," *Neural Computing & Applications*, vol. 14, no. 4, pp. 273–281, 2005.

[20] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28 – 44, 2013.

[21] 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.

[22] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, and J. Chanussot, "Challenges and opportunities of multimodality and data fusion in remote sensing," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, Sept 2014, pp. 106–110.

[23] K. Hedman, U. Stilla, G. Lisini, and P. Gamba, "Road network extraction in VHR SAR images of urban and suburban areas by means of class-aided feature-level fusion," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 3, pp. 1294–1296, March 2010.

[24] M. Pedergnana, P. Marpu, M. Dalla Mura, J. Benediktsson, and L. Bruzzone, "Classification of remote sensing optical and LiDAR data using extended attribute profiles," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 6, no. 7, pp. 856–865, Nov 2012.

[25] C. M. Bishop *et al.*, *Pattern recognition and machine learning*. springer New York, 2006, vol. 1.

[26] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRSS data-fusion contest," *IEEE Trans. Geosci. and Rem. Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct 2007.

[27] F. Pacifici, F. Del Frate, W. Emery, P. Gamba, and J. Chanussot, "Urban mapping using coarse SAR and optical data: Outcome of the 2007 GRSS data fusion contest," *IEEE Geosci. Rem. Sens. Lett.*, vol. 5, no. 3, pp. 331–335, July 2008.

[28] G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, C. Thiel, J. Inglada, E. Christophe, J. Chanussot, and P. Gamba, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRSS data fusion contest," *IEEE Trans. Geosci. and Rem. Sens.*, vol. 47, no. 11, pp. 3857–3865, Nov 2009.

[29] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot, and Q. Du, "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009 & 2010 data fusion contest," *IEEE JSTARS*, vol. 5, no. 1, pp. 331–342, Feb 2012.

[30] F. Pacifici, J. Chanussot, and Q. Du, "2011 GRSS data fusion contest: Exploiting WorldView-2 multi-angular acquisitions," in *IEEE IGARSS 2011*, July 2011, pp. 1163–1166.

[31] "Foreword to the special issue on optical multiangular data exploitation and outcome of the 2011 GRSS data fusion contest," *IEEE JSTARS*, vol. 5, no. 1, pp. 3–7, Feb 2012.

[32] F. Pacifici, Q. Du, and S. Prasad, "Report on the 2013 IEEE GRSS data fusion contest: Fusion of hyperspectral and LiDAR data [technical committees]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 3, pp. 36–38, Sept 2013.

[33] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pizurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici, "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 7, no. 6, pp. 2405–2418, June 2014.

[34] G. Moser, D. Tuia, and M. Shimoni, "2014 IEEE GRSS data fusion contest: Multiresolution fusion of thermal hyperspectral and VIS data [technical committees]," *Geoscience and Remote Sensing Magazine, IEEE*, vol. 2, no. 1, pp. 21–22, March 2014.

[35] R. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing, 3rd Ed.* Elsevier, 2007.

[36] C. Souza Jr, L. Firestone, L. M. Silva, and D. Roberts, "Mapping forest degradation in the eastern Amazon from SPOT 4 through spectral mixture models," *Remote Sens. Environ.*, vol. 87, no. 4, pp. 494–506, 2003.

[37] A. Mohammadzadeh, A. Tavakoli, V. Zoej, and J. Mohammad, "Road extraction based on fuzzy logic and mathematical morphology from pan-sharpened IKONOS images," *Photogramm. Rec.*, vol. 21, no. 113, pp. 44–60, 2006.

[38] F. Laporterie-Déjean, H. de Boissezon, G. Flouzat, and M.-J. Lefèvre-Fonollosa, "Thematic and statistical evaluations of five panchromatic/multispectral fusion methods on simulated PLEIADES-HR images," *Inform. Fusion*, vol. 6, no. 3, pp. 193–212, 2005.

[39] I. Amro, J. Mateos, M. Vega, R. Molina, and A. Katsaggelos, "A survey of classical methods and new trends in pansharpening of multispectral images," *EURASIP Journal on Advances in Signal Processing*, no. 79, pp. 1–22, Sep 2011.

[40] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Twenty-five years of pansharpening: A critical review and new developments," in *Signal and Image Processing for Remote Sensing, 2nd Ed.*, C. Chen, Ed. CRC Press, 2012, pp. 533–548.

[41] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. and Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, MAy 2008.

[42] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2565–2586, May 2015.

[43] W. Carper, T. Lillesand, and R. Kiefer, "The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 3, pp. 459–467, Apr 1990.

[44] P. Chavez, S. Sides, and J. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, Mar 1991.

[45] C. Laben and B. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening, Eastman Kodak Company," Patent US 6 011 875, 2000.

[46] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr 1983.

[47] K. Amolins, Y. Zhang, and P. Dare, "Wavelet based image fusion techniques - an introduction, review and comparison," *ISPRS Journal Photogramm*, vol. 62, no. 4, pp. 249–263, Sep 2007.

[48] V. Shah, N. Younan, and R. King, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008.

[49] L. Wald, T. Ranchin, and Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogram. Eng. Remote Sensing*, vol. 63, no. 6, pp. 691–699, Jun 1997.

[50] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb 2008.

[51] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. of International Conference on Image Processing (ICIP)*, vol. 2, 2003, pp. III – 173–1585.

[52] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *International journal of remote sensing*, vol. 10, no. 6, pp. 989–1003, 1989.

[53] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Review articledigital change detection methods in ecosystem monitoring: a review," *International journal of remote sensing*, vol. 25, no. 9, pp. 1565–1596, 2004.

[54] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *International journal of remote sensing*, vol. 25, no. 12, pp. 2365–2401, 2004.

[55] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 294–307, 2005.

[56] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 609–630, March 2013.

[57] B. Demir, F. Bovolo, and L. Bruzzone, "Detection of land-cover transitions in multitemporal remote sensing images with active-learning-based compound classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5, pp. 1930–1941, 2012.

[58] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 1, pp. 218–236, 2007.

[59] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2196–2212, 2012.

[60] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 1, pp. 4–19, 2003.

[61] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi-and hyperspectral data," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 463–478, 2007.

[62] M. J. Canty and A. A. Nielsen, "Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation," *Remote Sensing of Environment*, vol. 112, no. 3, pp. 1025–1036, 2008.

[63] G. Felde, G. Anderson, T. Cooley, M. Matthew, A. Berk, J. Lee *et al.*, "Analysis of hyperion data with the flaash atmospheric correction algorithm," in *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International*, vol. 1. IEEE, 2003, pp. 90–92.

[64] E. Karpouzli and T. Malthus, "The empirical line method for the atmospheric correction of IKONOS imagery," *International Journal of Remote Sensing*, vol. 24, no. 5, pp. 1143–1150, 2003.

[65] M. Cui, S. Prasad, W. Li, and L. M. Bruce, "Locality preserving genetic algorithms for spatial-spectral hyperspectral image classification," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 6, no. 3, pp. 1688–1697, 2013.

[66] D. Korycinski, M. M. Crawford, and J. W. Barnes, "Adaptive feature selection for hyperspectral data analysis," in *Remote Sensing*. International Society for Optics and Photonics, 2004, pp. 213–225.

[67] S. Prasad and L. M. Bruce, "Limitations of principal components analysis for hyperspectral target recognition," *Geoscience and Remote Sensing Letters, IEEE*, vol. 5, no. 4, pp. 625–629, 2008.

[68] D. Lunga, S. Prasad, M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 55–66, 2014.

[69] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 4, pp. 1185–1198, 2012.

[70] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 10, pp. 3973–3985, 2011.

[71] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 45–54, 2014.

[72] 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, IEEE*, vol. 3, no. 1, pp. 93–97, 2006.

[73] Y. Zhang and S. Prasad, "Locality preserving composite kernel feature extraction for multi-source geospatial image analysis," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 2014, to appear.

[74] Y. Zhang, H. Yang, S. Prasad, E. Pasolli, J. Jung, and M. Crawford, "Ensemble multiple kernel active learning for classification of multi-source remote sensing data," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 2014, to appear.

[75] H. Wu and S. Prasad, "Infinite Gaussian mixture models for robust decision fusion of hyperspectral imagery and full waveform LiDAR data," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*. IEEE, 2013, pp. 1025–1028.

[76] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio, "An empirical evaluation of deep architectures on problems with many factors of variation," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 473–480.

[77] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle *et al.*, "Greedy layer-wise training of deep networks," *Advances in neural information processing systems*, vol. 19, p. 153, 2007.

[78] N. Yokoya, S. Nakazawa, T. Matsuki, and A. Iwasaki, "Fusion of hyperspectral and LiDAR data for landscape visual quality assessment," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2419–2425, June 2014.

[79] D. Bar and S. Raboy, "Moving car detection and spectral restoration in a single satellite WorldView-2 imagery," *IEEE JSTARS*, vol. 6, no. 5, pp. 2077–2087, Oct 2013.

[80] F. Gao, B. Li, Q. Xu, and C. Zhong, "Moving vehicle information extraction from single-pass WorldView-2 imagery based on ERGAS-SNS analysis," *Remote Sensing*, vol. 6, no. 7, pp. 6500–6523, 2014.

[81] G. Licciardi, A. Villa, M. Dalla Mura, L. Bruzzone, J. Chanussot, and J. Benediktsson, "Retrieval of the height of buildings from WorldView-2 multi-angular imagery using attribute filters and geometric invariant moments," *IEEE JSTARS*, vol. 5, no. 1, pp. 71–79, Feb 2012.

[82] J. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *Geoscience and Remote Sensing Magazine, IEEE*, vol. 1, no. 2, pp. 6–36, June 2013.

[83] M. Daniel and A. Willsky, "A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 164–180, Jan 1997.

[84] J. A. Benediktsson, J. Chanussot, and W. M. Moon, "Very high-resolution remote sensing: challenges and opportunities," *Proceedings of the IEEE*, vol. 100, no. 6, pp. 1907–1910, 2012.

[85] J. Esper, P. V. Panetta, M. Ryschkewitsch, W. Wiscombe, and S. Neeck, "NASA-GSFC nano-satellite technology for Earth science missions," *Acta Astronautica*, vol. 46, no. 2, pp. 287–296, 2000.

[86] D. J. Barnhart, T. Vladimirova, A. M. Baker, and M. N. Sweeting, "A low-cost femtosatellite to enable distributed space missions," *Acta Astronautica*, vol. 64, no. 11, pp. 1123–1143, 2009.

[87] H. Ashida, K. Fujihashi, S. Inagawa, Y. Miura, K. Omagari, N. Miyashita, S. Matunaga, T. Toizumi, J. Kataoka, and N. Kawai, "Design of Tokyo Tech nano-satellite Cute-1.7+ APD II and its operation," *Acta Astronautica*, vol. 66, no. 9, pp. 1412–1424, 2010.

[88] P. Fahlstrom and T. Gleason, *Introduction to UAV systems*. John Wiley & Sons, 2012.

[89] S. Esterhuizen, "The design, construction, and testing of a modular gps bistatic radar software receiver for small platforms," Ph.D. dissertation, University of Colorado, 2006.

[90] K. Kotwal and S. Chaudhuri, "A novel approach to quantitative evaluation of hyperspectral image fusion techniques," *Information Fusion*, vol. 14, no. 1, pp. 5 – 18, 2013.