CHALLENGES AND OPPORTUNITIES OF MULTIMODALITY AND DATA FUSION IN REMOTE SENSING

M. Dalla Mura[1], *S. Prasad*[2], *F. Pacifici*[3], *P. Gamba*[4], *J. Chanussot*[1,5]

[1] GIPSA-Lab, Grenoble Institute of Technology, France
[2] University of Houston, USA
[3] DigitalGlobe Inc., CO, USA
[4] University of Pavia, Italy
[5] Faculty of Electrical and Computer Engineering, University of Iceland

ABSTRACT

Remote sensing is one of the most common ways to extract relevant information about the Earth through observations. Remote sensing acquisitions can be done by both active (SAR, LiDAR) and passive (optical and thermal range, multispectral and hyperspectral) devices. According to the sensor, diverse information of Earth's surface can be obtained. These devices provide information about the structure (optical, SAR), elevation (LiDAR) and material content (multi- and hyperspectral). Together they can provide information about land use (urban, climatic changes), natural disasters (floods, hurricanes, earthquakes), and potential exploitation (oil fields, minerals). In addition, images taken at different times can provide information about damages from floods, fires, seasonal changes etc. 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) which has been fostering the development of research and applications on this topic during the past decade.

Index Terms— Data fusion, remote sensing, pansharpening, classification, change detection

1. INTRODUCTION

The joint exploitation of different remote sensing sources is a key aspect towards a detailed characterization of the Earth. By focusing on the Earth's surface, remote sensing devices can be used for observing different aspects of the landscape, such as the spatial organization of objects in the scene, their height, identification of the constituent materials, characteristics of the material surfaces, composition of the underground, etc. Typically, a remote sensing device can only observe one or few of the aforementioned characteristics. Thus, in order to achieve a richer description of the scene, the observations derived by different acquisition sources should be coupled and jointly analyzed by *data fusion*.

In order to foster the research on this important topic, 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 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.

In this paper, we will present the opportunities and challenges of multimodal data fusion in remote sensing in the light of the outcomes of eight Data Fusion contests. The contests are presented in Section 2 and their outcomes in Section 3. Section 4 proposes a discussion of the opportunities and challenges of data fusion in remote sensing and Section 5 concludes this paper.

2. THE DATA FUSION CONTESTS

In the following, each Data Fusion Contest is briefly introduced by presenting the data used and the target application.

Data Fusion Contest 2006 [1] The focus of the 2006 Contest was on the fusion of multispectral and panchromatic images. Multispectral and panchromatic images are optical images acquired simultaneously by the same satellite (e.g., QuickBird) characterized by complementary characteristics

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

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

in terms of spatial resolution and number of spectral bands acquired. Multispectral images are typically characterized by a low spatial resolution and are composed by four spectral bands sensing the electromagnetic spectrum in four adjacent narrow intervals. The panchromatic image presents a greater spatial resolution (i.e., can better resolve spatial details of the surveyed scene) but it is monochromatic (i.e., it has a single spectral band). This data fusion problem aiming at producing a synthetic image with the spatial resolution of the panchromatic and the spectral resolution of the multispectral is referred to as *pansharpening* in the remote sensing community.

A set of simulated images from the Pleiades sensor and a spatially downsampled QuickBird image were provided to the participants. Each data set included very high spatial resolution (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 image was simulated in the Pleiades data set and it was the original multispectral image in the QuickBird one.

The results of the algorithms provided by the different research groups were compared with a standardized evaluation procedure, including both visual and quantitative analysis.

Data Fusion Contest 2007 [2] In the past two decades, monitoring urban centers and their peripheries at a regional scale has become an increasingly relevant topic for public institutions to keep track of the loss of agricultural land and natural vegetation due to urban development.
In 2007, the contest was related to the supervised classification of an urban area using both Synthetic Aperture Radar (SAR) and multispectral optical data. Multispectral sensors provide information about the energy reflected and radiated by the Earth's surface at different wavelengths, whilst SAR sensors are active devices that illuminate a scene with microwave pulses and register the lag and intensity of the signals backscattered by the objects. These acquisitions are characterized by high returns from buildings in urban areas and low and very low values from vegetated areas and water bodies, respectively. Due to the different imaging mechanisms, the data obtained by optical and SAR sensors differ greatly and can provide different information of the surveyed scene. The differences in the data result in challenges for their fusion. For instance optical and SAR images register different data types (real and complex, respectively), the type of noise affecting the data follows different models (mainly additive for optical and multiplicative for SAR).

Data Fusion Contest 2008 [3] In 2008, the contest was dedicated to the classification of VHR hyperspectral data. Hyperspectral images are optical images characterized by a VHR resolution since they perform acquisitions in up to hundreds of narrow adjacent intervals in the electromagnetic spectrum. A hyperspectral data set was distributed to every participant, and the task was to obtain a classified map as accurate as possible with respect to the ground truth data, depicting land-cover and land-use classes. In this contest the fusion was performed by the DFTC on the classification maps submitted by the participants. In fact, at each submission, the five best maps were combined using majority voting and re-ranked according to their respective contribution to the fused result.

Data Fusion Contest 2009-2010 [4] Data fusion could be particularly useful 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.

In 2009-2010, the contest was issued to address this task. Specifically, by performing change detection using multi-temporal and multi-modal data. The two pairs of data sets made available to the participates were acquired before and after a flood event. The class "change" was the area flooded by the river and class "no change" was the ground that stayed dry. The optical and SAR images were provided by CNES. The participants were allowed to use a supervised (i.e., using the available label of some pixels as a priori information for guiding the analysis) or an unsupervised method with all the data, the optical data only, or the SAR data only.

Data Fusion Contest 2011 [5, 6] One of the conclusions of the 2009-2010 Contest was the recommendation to investigate the use of new, additional, input features in order to increase the accuracy of remote sensing applications. The 2011 Contest focused on the use of an emerging type of images: VHR multiangular data (i.e., a sequence of images acquired from the same sensor on the same scene from different angles). In this case, none of the challenges of multisensoral data fusion apply since the images are all acquired by the same sensor. However, here registration problems should be solved in order to find the match in the different images of the same areas in the scene.

The unique data set was composed of five acquisitions on the same scene, including both 16 bit panchromatic and 8-band multispectral images acquired by WorldView-2. The scene was collected within a three minute time frame by different elevation angles.
For the first time, the organizers decided to keep the goal of the contest open to different research topics in order to focus on the different ways to exploit this kind of data. The various submissions presented the joint analysis of the multi-angular dataset for addressing tasks such as vegetation property retrieval, digital surface model and 3D building reconstruction, land-cover/land-use classification, the generation of a synthetic image with spatial resolution higher with respect to the acquired images (i.e., image super-resolution), object tracking, etc.

Data Fusion Contest 2012 [7] The 2012 Contest was designed to investigate the potential joint use of VHR multi-modal and multi-temporal images for various remote sensing applications. The data sets of this contest were acquired over the downtown San Francisco area, covering a number of large buildings, skyscrapers, commercial and industrial structures, a mixture of community parks and private housing, as well as highways and bridges. Three different types of data sets were provided to the participants, including spaceborne multispectral (i.e., QuickBird and WorldView-2 satellites) and VHR SAR imagery (acquired by TerraSAR-X), and airborne LiDAR data. The optical images show the scene as acquired by different sensors (i.e., having different spatial and spectral resolutions), viewing angles (i.e., leading to registration issues) and dates (i.e., acquisitions done under different conditions). The SAR images appear very different form the optical data since their metrical spatial resolution and the acquisition done over a dense urban area led to severe double bounce effects, layover and shadowing in the images. The LiDAR data report punctual information on the surface height.

Due to the great heterogeneity of this data set no specific application was targeted by the contest. Participants were asked to submit a paper describing the problem addressed, the method used, and the final results. Several contributions were received, the large majority of which investigated the fusion problem for urban land cover classification and change detection, followed by image pansharpening. Other topics included automated road extraction, moving object detection, urban tree inventory, and image super-resolution, demonstrating the large variety of applications that multi-modal/multi-temporal remote sensing images can offer.

Data Fusion Contest 2013 [8] The 2013 contest aimed at exploring the synergetic use of hyperspectral and LiDAR data for land cover classification. The hyperspectral imagery was composed of 144 spectral bands from 380 to 1050 nm. A co-registered Digital Surface Model (i.e., height map) derived from LiDAR data was also made available to all participants. Both data sets had the same spatial resolution (2.5 m). The challenge in this contest was the design of a classification procedure taking full advantage of these two data sources.

3. RESULTS OF THE CONTESTS

As can be seen in the previous section, these data fusion contests have covered a wide gamut, providing an insight on novel use of multi-modality data fusion for (1) Applications where similar modalities are used (e.g., passive optical) to create enhanced data products — specifically, pansharpened images; (2) Applications where disparate modalities (e.g., optical, SAR and LiDAR) are used simultaneously to better characterize the scene — be they problems related to traditional mapping via classification using multiple modalities, change detection between multi-temporal image sequences etc.. Additionally, these contests have attracted novel contributions "outside" the scope identified by the main challenge targeted by each contest. We describe some representative contributions along these directions below.

3.1. Pan-sharpening

In contest submissions pertaining to pan sharpening algorithms, quantitative results were possible owing to the available of reference originals (obtained either by simulating the data collected from the satellite sensor by means of high resolution data, or by first degrading all available data to a coarser resolution and using the original as reference). Simultaneously, visual analysis of the pan sharpened images was undertaken, keeping in mind radiometric and geometric qualities, focusing on linear features, punctual objects, surfaces, edges of buildings, roads, or bridges. Among the algorithms submitted to the contest, those taking into account the characteristics of the sensors (namely the modulation transfer function – i.e., the transfer function of the optical system) in order to perform the fusion proved to achieve the best results [1].

3.2. Change detection

The 2009-2010 contest data included optical and SAR imagery with the goal of change detection over a flooded area [4]. A variety of supervised and unsupervised approaches were proposed. 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 jointly SAR and optical data in an unsupervised scheme led to degraded performances with respect to the use of the sole optical data.

3.3. Classification

Various past contests have focused on the fusion of data in order to provide superior classification accuracy (with respect to considering the single modalities) for remote sensing applications. 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. 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.

3.4. Additional novel directions — beyond the challenges targeted by the contests

After the 2011 contest, the organizers have been encouraging submission on different research topics with respect to the main application addressed by the contest. This has resulted in very creative "out-of-the-box" thinking on new techniques to best combine multi-sensoral datasets. For instance, these submissions illustrated techniques to track moving objects (using either single or multiple in-track collections [9]), to retrieve building height [10], to derive new feature fusion methods based on graph theory, to applications such as visual quality assessment and modeling of thermal characteristics in urban environments. This has led to the emergence of new ideas that were not even envisioned when designing the contest. Other contributions proposed a method to derive an urban surface material map to parameterize a 3-dimensional numerical microclimate model. 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 [8].

4. DISCUSSION

As seen from the challenges proposed, data fusion can take place at different levels in the generic scheme aiming at extracting information from data. We can identify three main levels according to the type of outputs generated by the fusion.

- **Raw data level**. In some scenarios, the fusion of different modalities occurs 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. Pansharpening, super resolution and 3D reconstruction from 2D views are examples of applications that share this aim. This process can take advantage of some constraints that bound the problems such as the sensor used is the same (as when dealing with the 3D reconstruction), or as in pansharpening the different sensors are mounted on the same platform (not requiring a registration phase).
- **Feature level**. We refer to a fusion at the feature level when multisensoral data can be seen as an augmented set of observations, which can be taken jointly as input to a subsequent decision step. Focusing on classification, the simplest way to perform this fusion is to stack one type of data to the other and to feed the classifier with this new data set. In this case, the differences between the modes should be taken into accounts in order to be able to use them. For example, in the context of classification with Lidar and optical images, if one wants to use both the 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**. In this third case, the combination of the information coming from the different sources is done on the results obtained considering each modality separately. If the data provide complementary information for the application considered, through the fusion of the results obtained from each modality independently one can expect to increase the robustness of the decision. 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. An example of this type of fusion was presented in the 2008 [3] and 2009-10 [4] contests in which it was shown that the best results were obtained by fusing the best individual results.

For certain applications, the exploitation of multiple modalities 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 with a single acquisition. As for another example, 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.

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 not 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 in common formats for processing them jointly might not be straightforward. The fusion must be performed taking into account the characteristics of the sensors. 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 in 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.

Using multiple modalities does not always lead to improvements with respect to the use of a single mode (e.g., in the 2009-10 contest considering in an unsupervised scheme both optical and SAR images led to worse results than with respect to the sole optical). Indeed, considering data that are not relevant for the application could pollute 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.

5. CONCLUSION

By reviewing the outcomes of the last eight contests issued by the DFTC, we can start our concluding remarks by acknowledging the success of such initiative. These challenges have helped in catalyzing the research activity of the community on the broad field of data fusion in remote sensing. Specifically, the main contributions of the contests can be identified in i) fostering the methodological development on the topics defined by each contest; ii) making datasets available to the community — sometimes such datasets were kind of unique due to their rareness in real operative scenarios, such as for the contests 2012 (VHR SAR, VHR optical from different sensors and LiDAR) and 2013 (images VHR multiangular) and iii) encouraging the emergence of new applications or research directions based the data of the contests. From reviewing the different aspects of data fusion in remote sensing through the lens of the contests, one can clearly state that data fusion is indeed a promising way in order to extract information. Indeed, for some tasks and application it is the sole mean to perform the analysis. Nevertheless, 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. Towards that end, we look forward to the outcomes of the 2014 contest.

REFERENCES

[1] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L.M. 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.

[2] F. Pacifici, F. Del Frate, W.J. 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.

[3] 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.

[4] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot, and Qian 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.

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

[6] “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.

[7] 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.

[8] 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.

[9] D.E. 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.

[10] G.A. Licciardi, A. Villa, M. Dalla Mura, L. Bruzzone, J. Chanussot, and J.A. 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.