Processing of Extremely High-Resolution LiDAR and RGB Data: Outcome of the 2015 IEEE GRSS Data Fusion Contest–Part A: 2-D Contest

Manuel Campos-Taberner, Adriana Romero-Soriano, Carlo Gatta, Gustau Camps-Valls, *Senior Member, IEEE*, Adrien Lagrange, Bertrand Le Saux, Anne Beaupère, Alexandre Boulch, Adrien Chan-Hon-Tong, Stephane Herbin, Hicham Randrianarivo, Marin Ferecatu, Michal Shimoni, *Member, IEEE*, Gabriele Moser, *Senior Member, IEEE*, and Devis Tuia, *Senior Member, IEEE*

***Abstract*—In this paper, we discuss the scientific outcomes of the 2015 data fusion contest organized by the Image Analysis and Data Fusion Technical Committee (IADF TC) of the IEEE Geoscience and Remote Sensing Society (IEEE GRSS). As for previous years, the IADF TC organized a data fusion contest aiming at fostering new ideas and solutions for multisource studies. The 2015 edition of the contest proposed a multiresolution and multisensorial challenge involving extremely high-resolution RGB images and a three-dimensional (3-D) LiDAR point cloud. The competition was framed in two parallel tracks, considering 2-D and 3-D products, respectively. In this paper, we discuss the scientific results obtained by the winners of the 2-D contest, which studied either the complementarity of RGB and LiDAR with deep neural networks (winning team) or provided a comprehensive benchmarking evaluation of new classification strategies for extremely high-resolution multimodal data (runner-up team). The data and the previously undisclosed ground truth will remain available for the community and can be obtained at http://www.grss-ieee.org/community/technical-committees/data-fusion/2015-ieee-grss-data-fusion-contest/. The 3-D part of the contest is discussed in the Part-B paper [1].**

***Index Terms*—Deep neural networks, extremely high spatial resolution, image analysis and data fusion (IADF), landcover classification, LiDAR, multiresolution-, multisource-, multimodal-data fusion.**

I. Introduction to the 2015 Contest

The current development of Earth observation (EO) technologies, encompassing satellite missions, airborne acquisitions, drones, and unmanned aerial vehicles (UAV) is providing remote sensing scientists and practitioners with more and more opportunities to collect data of the Earth's surface for multiple global, regional, and local applications. These data can differ substantially in their physical natures (e.g., optical, thermal, radar, or laser observations), spatial resolutions (from a few centimeters to some kilometers using aerial and geostationary platforms, respectively), spectral resolutions (from panchromatic to hyperspectral imagery), and temporal resolutions (from a few minutes with geostationary systems to hours or days with constellations of near-polar satellites, and to on-demand acquisition with UAVs) [2].

In this framework, the capability to jointly benefit from those images critically depends on the development of accurate data fusion algorithms that effectively model the complementary information conveyed by distinct data sources [3]–[5]. Multisensor [6]–[8], multitemporal [9]–[11], and multiresolution [12]–[14] fusion techniques for remote sensing data have been researched for long, and are currently more and more relevant as they need to keep pace with the opportunities provided by these new data and the methodological challenges they raise [5]. It is in this framework that the IEEE Geoscience and Remote Sensing Society (IEEE GRSS) Image Analysis and Data Fusion Technical Committee (IADF TC[1]) organizes an annual Data Fusion Contest, in which a dataset is released free of charge to the international community along with a data fusion competition [7], [9], [12], [15]–[18]. This paper is the first of a two-part manuscript that aims at presenting and critically discussing the scientific outcomes of the 2015 edition of the Contest.

The 2015 Contest released to the international community of remote sensing an image dataset involving multiresolution

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

The work of A. Romero was supported by an APIF-UB grant. The work of C. Gatta was supported by MICINN under a Ramon y Cajal Fellowship. The work of M. Campos was supported by the European Union Seventh Framework Program FP7/2007-2013 under Grant 606983, and the LSA SAF (EUMETSAT) project. The work of G. Camps-Valls was supported by the European Research Council funding of the ERC-CoG-2014 SEDAL Consolidator Grant 647423, and the Spanish Ministry of Economy and Competitiveness for the funding through the Project LIFE-VISION TIN2012-38102-C03-01. The work of D. Tuia was supported by the Swiss National Science Foundation under Grant PP00P2-150593. (*Corresponding author: Devis Tuia.*)

M. Campos-Taberner and G. Camps-Valls are with the Universitat de València, Valencia 46010, Spain (e-mail: manuel.campos@uv.es; gcamps@uv.es).

A. Romero-Soriano is with the Universitat de Barcelona, Barcelona 08007 Spain (e-mail: adriana.romero@ub.edu).

C. Gatta is with the Universitat Autònoma de Barcelona, Barcelona 08193 Spain (e-mail: cgatta@cvc.uab.es).

A. Lagrange, B. Le Saux, A. Beaupére, A. Boulch, A. Chan-Hon-Tong, S. Herbin, and H. Randrianarivo are with the Office National d'Etudes et de Recherches Aérospatiales—The French Aerospace Lab, Palaiseau 91123, France (e-mail: adrien.lagrange@onera.fr; bertrand.le_saux@onera.fr; anne.beaupere@onera.fr; alexandre.boulch@onera.fr; adrien.chan_hon_tong@onera.fr; stephane.herbin@onera.fr; hicham.randrianarivo@onera.fr).

M. Ferecatu is with the Conservatoire National des Arts et Metiers – Cedric, Paris 75141, France (e-mail: Marin.Ferecatu@cnam.fr).

M. Shimoni is with the Signal and Image Centre, Department of Electrical Engineering, Royal Military Academy, Brussels 1000, Belgium (e-mail: mshimoni@elec.rma.ac.be).

G. Moser is with the Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture, University of Genoa, Genoa 16145, Italy (e-mail: gabriele.moser@unige.it).

D. Tuia is with the Department of Geography, University of Zurich, Zurich 8057, Switzerland (e-mail: devis.tuia@geo.uzh.ch).

[34] H. Sun, X. Sun, H. Wang, Y. Li, and X. Li, "Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 1, pp. 109–113, Jan. 2012.

[35] A. Cheriyadat, "Unsupervised feature learning for aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 439–451, Jan. 2014.

[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inform. Process. Syst.*, 2012, pp. 1097–1105.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–14.

[38] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[39] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proc. Adv. Neural Inform. Process. Syst.*, 2006, pp. 153–160.

[40] A. Coates and A. Ng, "The importance of encoding versus training with sparse coding and vector quantization," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 921–928.

[41] T. Schaul, S. Zhang, and Y. LeCun, "No more pesky learning rates," presented at Int. Conf. Machine Learning, Atlanta, GA, USA, 2013.

[42] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol: 54, no. 3, pp. 1349–1362, 2016.

[43] D. Davies and D. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 2, pp. 224–227, Apr. 1979.

[44] J. Dunn, "Well-separated clusters and optimal fuzzy partitions," *J. Cybern.*, vol. 4, no. 1, pp. 95–104, 1974.

[45] R. B. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Statist.*, vol. 3, pp. 1–27, 1974.

[46] J. Leitloff, S. Hinz, and U. Stilla, "Vehicle detection in very high resolution satellite images of city areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2795–2806, Jul. 2011.

[47] H. Randrianarivo, B. Le Saux, and M. Ferecatu, "Man-made structure detection with deformable part-based models," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2013, pp. 200–203.

[48] M. Fauvel, J. Chanussot, and J. Benediktsson, "Decision fusion for the classification of urban remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2828–2838, Oct. 2006.

[49] D. Tuia, F. Pacifici, M. Kanevski, and W. Emery, "Classification of very high spatial resolution imagery using mathematical morphology and support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3866–3879, Nov. 2009.

[50] V. Mnih and G. Hinton, "Learning to detect roads in high-resolution aerial images," in *Proc. Eur. Conf. Comp. Vis.*, 2010, pp. 210–223.

[51] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction of hyperspectral images," in *Proc. WHISPERS*, 2014.

[52] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comp. Vis.*, vol. 59, no. 2, pp. 167–181, 2004.

[53] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[54] H. Randrianarivo, B. Le Saux, M. Crucianu, and M. Ferecatu, "Discriminatively-trained model mixture for object detection in aerial images," presented at the Image Information Mining, Bucharest, Romania, 2015.

[55] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Co. Multimedia*, pp. 675–678, 2014.

[56] J. Yosinski, J. Clune, G. Hinton, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inform. Process. Syst.*, 2014, pp. 3320–3328.

[57] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inform. Process. Syst.*, 2012, pp. 1097–1105.

[58] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. Comput. Vis. Pattern Recog.*, 2009, pp. 248–255.

[59] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *Proc. Brit. Mach. Vis. Conf.*, 2014.

[60] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," in *Proc. Int. Conf. Learn. Rep.*, 2014. [Online]. Available: http://openreview.net/document/d332e77d-459a-4af8-b3ed-55ba

[61] A. Mahendran and A. Vedaldi, "Understanding deep image representations by inverting them," presented at the Computer Vision Pattern Recognition, Boston, MA, USA, 2015.

[62] K. Lenc and A. Vedaldi, "Understanding image representations by measuring their equivariance and equivalence," presented at the Computer Vision Pattern Recognition, Boston, MA, USA, 2015.

[63] S. Workman and N. Jacobs, "On the location dependence of convolutional neural network features," in *Proc. Computer Vision Pattern Recognition Worshops: EarthVision*, Jun. 2015, pp. 70–78.

[64] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proc. Comput. Vis. Pattern Recog. Worshops: EarthVision*, Jun. 2015, pp. 44–51.

[65] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep-learning earth observation classification using ImageNet pre-trained networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 105–109, Jan. 2016.

[66] S. Paisitkriangkrai, J. Sherrah, P. Janney, and A. Van-Den Hengel, "Effective semantic pixel labelling with convolutional networks and conditional random fields," in *Proc. Comput. Vis. Pattern Recog. Worshops: EarthVision*, Jun. 2015, pp. 36–43.