AN OVERVIEW OF MULTIMODAL REMOTE SENSING DATA FUSION: FROM IMAGE TO FEATURE, FROM SHALLOW TO DEEP

Danfeng Hong[1], *Jocelyn Chanussot*[2], *Xiao Xiang Zhu*[1,3]

[1]Remote Sensing Technology Institute, German Aerospace Center, Wessling, Germany
[2] Univ. Grenoble Alpes, INRIA, CNRS, Grenoble INP, LJK, Grenoble, France
[3] Data Science in Earth Observation, Technical University of Munich, Munich, Germany.

IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9554255

ABSTRACT

With the ever-growing availability of different remote sensing (RS) products from both satellite and airborne platforms, simultaneous processing and interpretation of multimodal RS data have shown increasing significance in the RS field. Different resolutions, contexts, and sensors of multimodal RS data enable the identification and recognition of the materials lying on the earth's surface at a more accurate level by describing the same object from different points of the view. As a result, the topic on multimodal RS data fusion has gradually emerged as a hotspot research direction in recent years.

This paper aims at presenting an overview of multimodal RS data fusion in several mainstream applications, which can be roughly categorized by 1) image pansharpening, 2) hyperspectral and multispectral image fusion, 3) multimodal feature learning, and (4) crossmodal feature learning. For each topic, we will briefly describe what is the to-be-addressed research problem related to multimodal RS data fusion and give the representative and state-of-the-art models from shallow to deep perspectives.

Index Terms— Classification, crossmodal, data fusion, deep learning, feature learning, multimodal, pansharpening, remote sensing, shallow models.

1. INTRODUCTION

Single remote sensing (RS) data, e.g., hyperspectral (HS) [1], Synthetic Aperture Radar (SAR) [2], Light Detection and Ranging (LiDAR) [3], and multispectral (MS) [4], inevitably meets the performance bottleneck in classifying the materials of interest, due to the lack of diverse information. With the ever-growing availability of different RS modalities, a variety of applications, such as land cover land use mapping [5, 6], mineral exploration [7, 8], object/target detection [9, 10], environmental monitoring [11], disaster response and management [12], have achieved a big performance improvement compared to that using only single modalities.

To provide a big picture of multimodal RS data fusion to the researchers, Ph.D. students, and senior engineers who are interested and would like to go deeper along this direction, we summarize some current state-of-the-art algorithms around this topic. This is further categorized by four sub-topics: image pansharpening, HS-MS data fusion, multimodal feature learning, and crossmodal feature learning.

2. IMAGE PANSHARPENING

Image pansharpening, as the name suggests, aims to sharpen the spatial components of the image, e.g., edge, texture, geometric structure, etc. This task can be also seen as a special fusion case. That is, the low-resolution MS image, can be super-resolved by the means of a single-band high resolution panchromatic (PAN) image.

A crucial survey on the pansharpening methods has been made in [13], where the two main types are considered for pansharpening, i.e., Component substitution (CS) and multiresolution analysis (MRA). The former one can be performed by learning a spectral transformation function to substitute the components of the MS data with the PAN image. Some well-known CS-based approaches, such as the intensity–hue–saturation (IHS) [14], principal component analysis (PCA) [15], the Gram–Schmidt (GS) [16], high-fidelity CS [17], matting model [18], and model-based reduced-rank [19], have been developed well for the pansharpening task. The latter one aims to inject multiresolution detailed information of the PAN image into the resampled MS bands, yielding the high-resolution MS products. These multiresolution extraction algorithms consist of decimated wavelet transform (DWT) [20], undecimated wavelet transform (UDWT) [21], "a trous" wavelet transform (ATWT) [22], super-wavelets transforms(e.g., contourlet [23], curvelet [24]).

Beyond traditional pansharpening models, deep learning (DL) techniques, owing to its powerful learning ability, have achieved more potential results. A representative method presented in [25] to perform the task is PNN by the means of convolutional neural networks (CNNs), named PNN. Following this framework, numerous DL-based pansharpening networks, e.g., DRPNN [26], PanNet [27], MSDCNN [28], PNN+ [29], RSIFNN [30], pyramidNet [31], DiCNN [32], have been successively proposed to further enhance the spatial details of pan-sharpened images.

3. HYPERSPECTRAL AND MULTISPECTRAL IMAGE FUSION

Characterized by high spectral resolution, HS images enable us to detect the objects of interest more easily. However, there is a trade-off between the spatial and spectral resolutions of the HS image. For this reason, enormous effects have been made to enhance spatial resolution of HS images by fusing overlapped MS images with high resolution.

The subspace-based approach is a representative traditional model for fusing the HS-MS images by means of coupled matrix factorization or spectral unmixing [33, 34, 35] or Bayesian estimation [36, 37, 38, 39]. Beyond the 2-D modeling, the researchers have attempted to directly consider the original HS image as a 3-D tensor to simultaneously capture its spatial and spectral properties. As a result, Tucker and CP decomposition approaches are used to model the fusion problem, leading to a coupled sparse tensor factorization approach [40, 41].

Very recently, some preliminary works related to the HS-MS fusion task have been proposed by designing advanced deep networks [42, 43, 44, 45, 46, 47]. These models have demonstrated their effectiveness and superiority in the fusion task. However, some remaining challenges, e.g., lack of prior knowledge, limited training pairs, etc. still hinder the performance improvement.

4. MULTIMODAL FEATURE LEARNING

Different from the image level data fusion, multimodal feature learning (MFL) directly learns feature level fusion. Many popular approaches related to MFL have been largely proposed and widely applied in various high-level applications, such as classification, object detection, image segmentation.

Morphological profiles are a classic and well-known methodology to extract and fuse rich spatial information of multimodal RS images [48, 49, 50]. Another group is the manifold learning based approach. It aims at learning the shared representations by aligning different modalities of RS data, thereby achieving the information exchange [51, 52, 53].

Inspired by the recent success of DL techniques, some exploratory researches [54, 55, 56, 57] have been developed to perform the feature learning with the multimodal data input. It is worthy noting that a general and unified DL framework is proposed for multimodal RS image classification [58]. This is a seminal work in the RS community, which has been garnering increasing attention by researchers.

5. CROSSMODAL FEATURE LEARNING

It is well known that the overlapped or registered data from different sensors or platforms are expensive to be collected simultaneously, especially on a large-scale region. Therefore, how to only utilize partially overlapped multimodal data to process and analysis a large-scale RS data is a to-be-solved challenge [59]. This is a typical crossmodal feature learning (CFL) issue. More specifically, CFL is defined as "multiple modalities are used for training but several modalities are missing in the testing phase, or *vice versa* [60]".

Manifold alignment (MA) is an effective solution with respect to the CFL's issue, since it can align multiple modalities on manifold subspaces, further achieving information transfer and retrieval of different modalities [61]. Beyond MA-based approaches, shared subspace learning (SSL) directly bridges the learned features and label information, yielding more competitive fusion performance in features. In recent years, some SLL-related methods have been developed in a supervised or semisupevised way [62, 63, 6, 1].

Although a large number of deep networks have been designed to address this issue in computer vision or machine learning, yet it is still less investigated in RS. Until now, only a few DL-related works are made to provide possible and potential solutions for the CFL's issue, e.g., [59, 64].

6. CONCLUSION

In this paper, we provide an overview on multimodal RS data fusion from different perspectives, such as image level, feature level, shallow model, and deep model. More specifically, we briefly introduce four different research directions, they are image pansharpening, HS and MS image fusion, multimodal feature learning, and its special case: crossmodal feature learning. We believe that multimodal RS, particularly multimodal data processing and analysis, e.g., fusion, would be the leading-edge research hotspot in the future, which is worth paying close attention to.

7. ACKNOWLEDEGEMENT

This work is jointly supported by the German Research Foundation (DFG) under grant ZH 498/7-2, by the Helmholtz Association through the framework of Helmholtz Artificial Intelligence (HAICU) - Local Unit "Munich Unit @Aeronautics, Space and Transport (MASTr)". This work of J. Chanussot is supported by the MIAI@Grenoble Alpes (ANR-19-P3IA-0003) and the AXA Research Fund.

8. REFERENCES

[1] D. Hong, L. Gao, J. Yao, B. Zhang, P. Antonio, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, 2020. DOI: 10.1109/TGRS.2020.3015157.

[2] J. Chanussot, G. Mauris, and P. Lambert, "Fuzzy fusion techniques for linear features detection in multitemporal sar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1292–1305, 1999.

[3] R. Huang, D. Hong, Y. Xu, W. Yao, and U. Stilla, “Multi-scale local context embedding for lidar point cloud classification,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 721–725, 2019.

[4] D. Hong, W. Liu, J. Su, Z. Pan, and G. Wang, “A novel hierarchical approach for multispectral palmprint recognition,” *Neurocomputing*, vol. 151, pp. 511–521, 2015.

[5] E. Bartholome and A. S. Belward, “Glc2000: a new approach to global land cover mapping from earth observation data,” *Int. J. Remote Sens.*, vol. 26, no. 9, pp. 1959–1977, 2005.

[6] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. Zhu, “Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification,” *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 193–205, 2019.

[7] F. F. Sabins, “Remote sensing for mineral exploration,” *Ore Geol. Rev.*, vol. 14, no. 3-4, pp. 157–183, 1999.

[8] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, “An augmented linear mixing model to address spectral variability for hyperspectral unmixing,” *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, 2019.

[9] L. Gao, B. Yang, Q. Du, and B. Zhang, “Adjusted spectral matched filter for target detection in hyperspectral imagery,” *Remote Sens.*, vol. 7, no. 6, pp. 6611–6634, 2015.

[10] X. Wu, D. Hong, J. Chanussot, Y. Xu, R. Tao, and Y. Wang, “Fourier-based rotation-invariant feature boosting: An efficient framework for geospatial object detection,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 302–306, 2020.

[11] S. L. Ustin, *Manual of Remote Sensing, Remote Sensing for Natural Resource Management and Environmental Monitoring*, vol. 4. John Wiley & Sons, 2004.

[12] S. Voigt, T. Kemper, T. Riedlinger, R. Kiefl, K. Scholte, and H. Mehl, “Satellite image analysis for disaster and crisis-management support,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1520–1528, 2007.

[13] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, “A critical comparison among pansharpening algorithms,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, 2014.

[14] 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. 4, pp. 459–467, 1990.

[15] V. P. Shah, N. H. Younan, and R. L. 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, 2008.

[16] C. A. Laben and B. V. Brower, “Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening,” 2000. US Patent 6,011,875.

[17] Q. Xu, B. Li, Y. Zhang, and L. Ding, “High-fidelity component substitution pansharpening by the fitting of substitution data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7380–7392, 2014.

[18] X. Kang, S. Li, and J. A. Benediktsson, “Pansharpening with matting model,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5088–5099, 2013.

[19] F. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, “Model-based reduced-rank pansharpening,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 656–660, 2019.

[20] S. G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, 1989.

[21] G. P. Nason and B. W. Silverman, “The stationary wavelet transform and some statistical applications,” in *Wavelets and statistics*, pp. 281–299, Springer, 1995.

[22] M. J. Shensa, “The discrete wavelet transform: wedding the a trous and mallat algorithms,” *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2464–2482, 1992.

[23] M. N. Do and M. Vetterli, “The contourlet transform: an efficient directional multiresolution image representation,” *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, 2005.

[24] J. Ma and G. Plonka, “The curvelet transform,” *IEEE Signal Process. Mag.*, vol. 27, no. 2, pp. 118–133, 2010.

[25] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, “Pansharpening by convolutional neural networks,” *Remote Sens.*, vol. 8, no. 7, p. 594, 2016.

[26] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, “Boosting the accuracy of multispectral image pansharpening by learning a deep residual network,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, 2017.

[27] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, “Pannet: A deep network architecture for pan-sharpening,” in *Proc. ICCV*, pp. 5449–5457, 2017.

[28] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, “A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 3, pp. 978–989, 2018.

[29] G. Scarpa, S. Vitale, and D. Cozzolino, “Target-adaptive cnn-based pansharpening,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, 2018.

[30] Z. Shao and J. Cai, “Remote sensing image fusion with deep convolutional neural network,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, 2018.

[31] Y. Zhang, C. Liu, M. Sun, and Y. Ou, “Pan-sharpening using an efficient bidirectional pyramid network,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5549–5563, 2019.

[32] L. He, Y. Rao, J. Li, J. Chanussot, A. Plaza, J. Zhu, and B. Li, “Pansharpening via detail injection based convolutional neural networks,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 4, pp. 1188–1204, 2019.

[33] N. Yokoya, T. Yairi, and A. Iwasaki, “Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, 2011.

[34] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, “Hyperspectral and multispectral image fusion based on a sparse representation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, 2015.

[35] X. Li, Y. Yuan, and Q. Wang, “Hyperspectral and multispectral image fusion based on band simulation,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 3, pp. 479–483, 2019.

[36] Y. Zhang, S. De Backer, and P. Scheunders, “Noise-resistant wavelet-based bayesian fusion of multispectral and hyperspectral images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3834–3843, 2009.

[37] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *Proc. ICASSP*, pp. 3176–3180, IEEE, 2014.

[38] B. Lin, X. Tao, M. Xu, L. Dong, and J. Lu, "Bayesian hyperspectral and multispectral image fusions via double matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5666–5678, 2017.

[39] L. Sui, L. Li, J. Li, N. Chen, and Y. Jiao, "Fusion of hyperspectral and multispectral images based on a bayesian nonparametric approach," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 4, pp. 1205–1218, 2019.

[40] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, 2018.

[41] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6503–6517, 2018.

[42] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3d full convolutional neural network," *Remote Sens.*, vol. 9, no. 11, p. 1139, 2017.

[43] X. Liu, C. Deng, J. Chanussot, D. Hong, and B. Zhao, "Stfnet: A two-stream convolutional neural network for spatiotemporal image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6552–6564, 2019.

[44] W. Liu and J. Lee, "An efficient residual learning neural network for hyperspectral image superresolution," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 4, pp. 1240–1253, 2019.

[45] K. Zheng, L. Gao, W. Liao, D. Hong, B. Zhang, X. Cui, and J. Chanussot, "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2487–2502, 2020.

[46] T. Uezato, D. Hong, N. Yokoya, and W. He, "Guided deep decoder: Unsupervised image pair fusion," in *Proc. ECCV*, pp. 87–102, Springer, 2020.

[47] J. Yao, D. Hong, J. Chanussot, D. Meng, X. Zhu, and Z. Xu, "Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution," in *Proc. ECCV*, pp. 208–224, Springer, 2020.

[48] W. Liao, A. Pižurica, R. Bellens, S. Gautama, and W. Philips, "Generalized graph-based fusion of hyperspectral and lidar data using morphological features," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 552–556, 2014.

[49] N. Yokoya, P. Ghamisi, J. Xia, S. Sukhanov, R. Heremans, I. Tankoyeu, B. Bechtel, B. L. Saux, G. Moser, and D. Tuia, "Open data for global multimodal land use classification: Outcome of the 2017 ieee grss data fusion contest," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 5, pp. 1363–1377, 2018.

[50] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, 2020.

[51] D. Tuia, M. Volpi, M. Trolliet, and G. Camps-Valls, "Semisupervised manifold alignment of multimodal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7708–7720, 2014.

[52] J. Hu, D. Hong, and X. X. Zhu, "Mima: Mapper-induced manifold alignment for semi-supervised fusion of optical image and polarimetric sar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9025–9040, 2019.

[53] D. Hong, W. He, N. Yokoya, J. Yao, L. Gao, L. Zhang, J. Chanussot, and X. X. Zhu, "Interpretable hyperspectral artificial intelligence: When non-convex modeling meets hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Mag.*, 2021. DOI: 10.1109/MGRS.2021.3064051.

[54] P. Benedetti, D. Ienco, R. Gaetano, K. Ose, R. Pensa, and S. Dupuy, "M^3Fusion: A deep learning architecture for multiscale multimodal multitemporal satellite data fusion," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 12, pp. 4939–4949, 2018.

[55] R. Hang, Z. Li, P. Ghamisi, D. Hong, G. Xia, and Q. Liu, "Classification of hyperspectral and lidar data using coupled cnns," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4939–4950, 2020.

[56] D. Hong, L. Gao, R. Hang, B. Zhang, and J. Chanussot, "Deep encoder-decoder networks for classification of hyperspectral and lidar data," *IEEE Geosci. Remote Sens. Lett.*, 2020. DOI:10.1109/LGRS.2020.3017414.

[57] X. Wu, W. Li, D. Hong, J. Tian, R. Tao, and Q. Du, "Vehicle detection of multi-source remote sensing data using active fine-tuning network," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 39–53, 2020.

[58] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, 2021.

[59] D. Hong, N. Yokoya, G.-S. Xia, J. Chanussot, and X. X. Zhu, "X-modalnet: A semi-supervised deep cross-modal network for classification of remote sensing data," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 12–23, 2020.

[60] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Ng, "Multimodal deep learning," in *Proc. ICML*, pp. 689–696, 2011.

[61] C. Wang and S. Mahadevan, "A general framework for manifold alignment.," in *Proc. AAAI*, pp. 79–86, 2009.

[62] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, "CoSpace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geos. Remote Sens.*, vol. 57, no. 7, pp. 4349–4359, 2019.

[63] D. Hong, J. Chanussot, N. Yokoya, J. Kang, and X. X. Zhu, "Learning-shared cross-modality representation using multispectral-lidar and hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1470–1474, 2020.

[64] D. Hong, J. Yao, D. Meng, Z. Xu, and J. Chanussot, "Multimodal gans: Toward crossmodal hyperspectral-multispectral image segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, 2020. DOI: 10.1109/TGRS.2020.3020823.