Multimodal Data Fusion

By TÜLAY ADALI, *Fellow, IEEE*

CHRISTIAN JUTTEN, *Fellow, IEEE*

LARS KAI HANSEN

In many fields today, information about a given phenomenon is obtained through different types of acquisition techniques and experimental conditions, and the availability of such multimodal data has been growing. Joint analysis of this data—its fusion—promises a more comprehensive and informative view of the task at hand, and, if probed carefully, may open new venues and answer questions we might not have even thought of asking when working with a single modality.

This increasing availability of multimodal data has spurned significant activity in disciplines spanning medical imaging, remote sensing, speech processing, behavioral sciences, metabolomics, to name a few. Thus, ways to leverage the use of complementary information is of interest in numerous fields. While there are a number of challenges that are specific to a given domain, a good number of the challenges researchers face in such analyses are common to all. Despite this shared interest, however, there has been little communication, if any, among disciplines to allow cross fertilization of ideas and sharing of the common experience.

This special issue provides a review of current approaches and results on multimodal data fusion from different disciplines, and by including the required background in each domain, aims to provide a common forum for the exchange of ideas in this very active field of research.

This is the goal of our special issue, while providing a review of current approaches and results on multimodal data fusion from disciplines where there has been important activity, to also provide a forum for the exchange of ideas in this very active field of research. Hence, we have asked our contributors to make sure that they would introduce their field and the problems they are addressing clearly, in a manner accessible to those outside the field, and to specifically address a number of key challenges that are common across various disciplines. This is a long list and includes challenges at multiple levels such as at the data and the design levels. Among those, a few notable ones are the following:

- How do we work with data at different scales, resolutions and of different types?
- How do we best deal with noisy, uncertain, missing, and conflicting data?
- How do we best allow the interaction of multiple modalities, and at what level?
- How do we evaluate performance and how do we make sure that bringing in multiple modalities to the solution does actually improve the end result?
- How do we establish the link among the data sets, and when establishing the link, what are the most reasonable assumptions, especially given the unknown nature of interactions across the data sets?

We have gathered 11 papers that address these challenges and introduce the state of the art in disciplines ranging from remote sensing and medical imaging to chemometrics. The first paper, "Multimodal data

Digital Object Identifier: 10.1109/JPROC.2015.2461273

0018-9219 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

fusion: An overview of methods, challenges, and prospects" by Lahat *et al.*, sets the stage for all the others by providing an overview of the main challenges in multimodal data fusion and motivates those by numerous examples across disciplines. The paper deals with two key questions: "why we need data fusion" and "how we perform data fusion." To address the latter question, "diversity" is introduced as a key concept, and a number of data-driven solutions based on matrix and tensor decompositions are discussed emphasizing how they account for diversity across the data sets. As noted in this paper, data-driven methods arise naturally as the desirable solution for multimodal data fusion since usually very little is known about the interactions of different modalities, and very few *a priori* assumptions, if any, could be made in general. The next contribution, "Multimodal data fusion using source separation" by Adalı *et al.*, is in two parts and introduces two powerful data-driven models that aim to maximally exploit the statistical diversity within and across the data sets. The first paper in the sequence, "Multimodal data fusion using source separation: Two effective models based on ICA and IVA and their properties," explains the two models, discusses their properties, and gives implementation details to enable the user to make informed decisions on the choice of a given model and the associated parameters. The discussion in the paper does not center around a certain application to emphasize the general applicability of the models, and it is the second paper in the sequence, "Multimodal data fusion using source separation: Application to medical imaging," that demonstrates the application of these models to fusion of medical imaging data from three modalities, two functional and one structural modality—functional magnetic resonance imaging (MRI), structural MRI, and electroencephalography data. The paper discusses the tradeoffs in design for the two models and their performance for this practical problem. The difficult question of validation and interpretation of results is also addressed.

Fusion in medical image analysis is the continuing theme for the next two papers. In "Multivariate machine learning methods for fusing multimodal functional neuroimaging data," Dähne *et al.* review multivariate approaches used for the fusion of functional neuroimaging data. They focus on methods based on factor analysis and emphasize the physical interpretability of the resulting factor model parameters. They note the applicability of the methods to other problems, and for the given application domain, they address multiple advantages such as better neuroscientific understanding and enhanced diagnostic performance. In "Tensor analysis and fusion of multimodal brain images," Karahan *et al.* note the suitability of multiway (tensor) models to represent the multimodal and multiscale nature of the neuroimaging data and introduce the so-called Markov–Penrose diagrams for the illustration of such models and demonstrate that Granger causal analysis can be cast as a tensor regression problem.

Earth observation through satellite and airborne sensors provides heterogeneous images of the same scene. Multispectral, hyperspectral, radar, Lidar, multitemporal, and multiangular images can be available for the same geographical region. In "Multimodal classification of remote sensing images: A review and future directions," Gomez-Chova *et al.* provide an overview of approaches for combining this multimodal data to improve classification of the materials on the surface. While providing a thorough review of current methodologies in seven challenging data fusion applications for remote sensing, they highlight the promising methods that benefit from the synergy between machine learning and signal processing: sparse methods, kernel-based fusion, Markov modeling, and manifold alignment. The next paper by Dalla Mura *et al.*, "Challenges and opportunities of multimodality and data fusion in remote sensing," leverages the results of the data fusion contests organized by the IEEE Geoscience and Remote Sensing Society to address a number of key questions, such as "how multimodality is considered and integrated in the processing chain" and "what the actual gains have been." The results of the competition are used to motivate a comprehensive discussion on the main challenges and perspectives for data fusion in remote sensing as well as in other fields.

Metabolomics is a rich source of heterogeneous data sets, which can be represented as matrices as well as tensors, recorded with various techniques such as gas or liquid chromatography-mass spectrometry, nuclear magneticresonance, and fluorescence spectroscopy, among others. In this context, in the paper "Data fusion in metabolomics using coupled matrix and tensor factorizations," Acar *et al.* address the data fusion challenge by jointly factorizing matrix and tensor data sets for extracting shared as well as unshared factors. This is achieved by adding regularization terms within the objective function for modeling desirable relationships between factors. In the paper "Component- and factor-based models for data fusion in the behavioral sciences," Van Mechelen *et al.* consider multiblock and factor-based methods for data fusion on coupled data matrices, which all share the same mode. They especially focus on two closely related families of models for coupled data matrices, all of which share either the experimental unit or the variable data mode, and they show how to capture both similarities and differences in the data matrices by simply varying certain terms in the objective function.

Digital media provides an application domain of significant practical and theoretical importance. In speech recognition, for example, video provides valuable information as it is not affected by disturbances such as ambient noise and reverberations. In "Audiovisual fusion: Challenges and new approaches," Katsaggelos *et al.*

discuss main challenges in the area with a focus on desynchronization of the two modalities as well as the issue of training and testing where one of the modalities might be absent for testing. They discuss a number of solutions for the problem with the emphasis on multiview learning where this problem is addressed by learning a representation of the modalities by taking advantage of the multimodal data during the training phase. Finally, in "Three-dimensional imaging with multiple degrees of freedom using data fusion," Latorre-Carmona *et al.* introduce another dimension of data fusion, the use of different information sources such as polarimetry and multispectral imaging, with integral imaging and digital holographic systems to achieve 3-D imaging. They demonstrate that 3-D imaging using fusion provides multiple advantages, including enhancement in visualization and reconstruction under different acquisition conditions.

In closing, we thank our contributors for their comprehensive and stimulating articles and Vaishali Damle, Managing Editor of the Proceedings of the IEEE, for her guidance and support throughout the project. We would like to also extend our thanks to the reviewers for their detailed and insightful comments, and to Jo Sun, Senior Publications Editor, Proceedings of the IEEE, for her special care in putting together this special issue, and her amazingly prompt responses to our many requests during the process.

Multimodal data fusion, we believe, is one of the most exciting and active areas of research today. However, despite the great amount of activity, this is still an area very much in its infancy. Hence, all of the papers in our special issue, besides providing a review of the promising solutions in an array of disciplines, also keep the focus on the challenges and prospects multimodal data fusion presents. Indeed, while multimodality poses a number of challenges, it also offers the possibility to ask a rich set of questions we might not have thought of asking before. We hope the collection in our special issue will stimulate discussion in this exciting area of research, and more importantly, will allow cross fertilization of ideas across the many disciplines where multimodal data is now becoming increasingly common. ■

ABOUT THE GUEST EDITORS

Tülay Adalı (Fellow, IEEE) received the Ph.D. degree in electrical engineering from North Carolina State University, Raleigh, NC, USA, in 1992.

She joined the faculty at the University of Maryland Baltimore County (UMBC), Baltimore, MD, USA, in 1992. She is currently a Distinguished University Professor in the Department of Computer Science and Electrical Engineering at UMBC. Her research interests are in the areas of statistical signal processing, machine learning for signal processing, and biomedical data analysis.

Prof. Adalı assisted in the organization of a number of international conferences and workshops including the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), the IEEE International Workshop on Neural Networks for Signal Processing (NNSP), and the IEEE International Workshop on Machine Learning for Signal Processing (MLSP). She was the General Co-Chair, NNSP (2001–2003); Technical Chair, MLSP (2004–2008); Program Co-Chair, MLSP (2008, 2009, and 2014), 2009 International Conference on Independent Component Analysis and Source Separation; Publicity Chair, ICASSP (2000 and 2005); and Publications Co-Chair, ICASSP 2008. She is Technical Program Co-Chair for ICASSP 2017 and Special Sessions Co-Chair for ICASSP 2018. She chaired the IEEE Signal Processing Society (SPS) MLSP Technical Committee (2003–2005, 2011–2013), served on the SPS Conference Board (1998–2006), and the Bio Imaging and Signal Processing Technical Committee (2004–2007). She was an Associate Editor for the IEEE Transactions on Signal Processing (2003–2006), the IEEE Transactions on Biomedical Engineering (2007–2013), the IEEE Journal of Selected Areas in Signal Processing (2010–2013), and *Elsevier Signal Processing Journal* (2007–2010). She is currently serving on the Editorial Boards of the Proceedings of the IEEE and *Journal of Signal Processing Systems for Signal, Image, and Video Technology*, and is a member of the IEEE Signal Processing Theory and Methods Technical Committee. She is a Fellow of the American Institute of Biomedical and Medical Engineers (AIMBE), a Fulbright Scholar, recipient of a 2010 IEEE Signal Processing Society Best Paper Award, 2013 University System of Maryland Regents' Award for Research, and an NSF CAREER Award. She was an IEEE Signal Processing Society Distinguished Lecturer for 2012 and 2013.

Christian Jutten (Fellow, IEEE) received the Ph.D. and Doctor ès Sciences degrees in signal processing from Grenoble Institute of Technology (GIT), Grenoble, France, in 1981 and 1987, respectively.

From 1982, he was an Associate Professor at GIT, before becoming Full Professor at the University Joseph Fourier of Grenoble, Grenoble, France, in 1989. For 30 years, his research interests have been machine learning and source separation, including theory (separability, source separation in nonlinear mixtures, sparsity, multimodality) and applications (brain and hyperspectral imaging, chemical sensor array, speech). He is author or coauthor of more than 85 papers in international journals, four books, 25 keynote plenary talks, and 190 communications in international conferences. He has been Visiting Professor at Swiss Federal Polytechnic Institute (Lausanne, Switzerland, 1989), at Riken labs (Japan, 1996) and at Campinas University (Brazil, 2010). He was Director or Deputy Director of his lab from 1993 to 2010, as well as Head of the Signal Processing Department (120 people) and Deputy Director of GIPSA-lab (300 people) from 2007 to 2010. He was a scientific advisor for signal and images processing at the French Ministry of Research (1996–1998) and for the French National Research Center (2003–2006). Since May 2012, he has been Deputy Director at the Institute for Information Sciences, French National Center of Research (CNRS) in charge of signal and image processing.

Dr. Jutten was organizer or Program Chair of many international conferences, including the 1st International Conference on Blind Signal Separation and Independent Component Analysis in 1999. He was a member of a few IEEE Technical Committees, and currently is a member of "Theory and Methods" of the IEEE Signal Processing Society. He received best paper awards from the European Association for Signal Processing (EURASIP) in 1992 and the IEEE Geoscience and Remote Sensing Society (GRSS) in 2012, and the 1997 Medal Blondel from the French Electrical Engineering Society for his contributions in source separation and independent component analysis. He is the EURASIP Fellow (since 2013). He has been a Senior Member of the Institut Universitaire de France since 2008, with renewal in 2013. He is the recipient of a 2012 ERC Advanced Grant for a project on challenges in extraction and separation of sources (CHESS).

Lars Kai Hansen received the Ph.D. degree in physics from the University of Copenhagen, Copenhagen, Denmark, in 1986.

Since 2000, he has held the professorship in digital signal processing at the Technical University of Denmark, Kgs. Lyngby, Denmark, where he also heads the Cognitive Systems Section. His research focuses on adaptive signal processing and machine learning with applications in biomedicine, digital media, and cognitive systems. He has published more than 300 papers on these subjects in journals, conferences, and books. The major contributions include introduction and analysis of neural network ensembles (1990–1996); the first application of predictive machine learning for mind reading in PET (1994–present) and fMRI (1997–present); the concept of cognitive components (2005–present); the cure for variance inflation (2001, 2011–present). His research is generously supported by Danish and international research councils and foundations, including the U.S. National Institutes of Health, the European Union, the Danish Innovation Foundation, and the Lundbeck Foundation. In 2011, he served as Cátedra de Excelencia at Universidad Carlos III de Madrid.