Introduction to the Special Section on Image Information Mining for Earth Observation Data

EARTH observation (EO) data have increased significantly over the last decades with orbital and suborbital sensors collecting and transmitting to Earth receiving stations several terabytes of data a day. This data acquisition rate is a major challenge to the existing data exploitation and dissemination approaches used by the various agencies (e.g., ESA, NASA, NOAA, or national agencies) charged with extracting information from these images. And, with plans for more EO systems (NPOESS, ENVISAT, TerraSAR-X, GMES, and commercial systems) the challenge is increasingly going to be how to enlarge the usability of the millions of images being stored in archives to a larger and larger group of end-user applications (e.g., climate change, security, land use, weather). To help support knowledge discovery from EO images and archives, researchers around the world have begun to tackle the formidable challenge of developing concepts, tools, and applications for extracting information from the petabytes of EO images archived globally.

Image information mining (IIM) is a new field of study that has arisen to seek solutions to automating the mining (extracting) of information from EO archives that can lead to knowledge discovery and the creation of actionable intelligence (exploiting). Image information mining is more than just an extension of data mining principles to images. Also, the mining of EO images for content information is different from mining images of facial characteristics or of other living things (e.g., lions or tigers) due to the fact there are no features (e.g., ears, stripes, or wings) that have known relationships to help differentiate classes in an EO scene. Therefore, IIM is an interdisciplinary approach to automating remote sensing analysis that draws on signal/image analysis, pattern recognition, artificial intelligence, machine learning, information theory, databases, semantics, ontologies, and knowledge management.

IIM includes novel concepts and methods to help humans to access and discover information in large image archives, to rapidly gather information about courses of action. Interesting applications involve complicated spatial, structural, and temporal relationships among image objects. Thus, new concepts have been introduced on the basis of intensive preprocessing of images to extract relevant features, structures, and objects and automatically record and analyze their interrelationships to learn their behavior so as to be able to detect relevant information. The methods are integrated in systems, which can be operated using intelligent interfaces able to correlate the information content of the images with the relevant goals of the application. The users have at their disposition tools for the definition of specific goals using semantics. The problem of the large dimensionality, which for computationally efficient data analysis is of primary concern, is solved using preextracted representative features instead of raw images. These are difficult tasks which require cooperative solutions integrating a variety of methods of soft computing, information semantics and the semantic web, advanced statistics, and probabilistic reasoning. The goal is to have machines more closely interacting at human conceptual levels (i.e., automate the human remote sensing analyst). Unlike the respective hard computing methods, soft computing may cope with problems that deal with imprecision, uncertainty, and learning, and are better candidates to construct systems and models which are simple, applicable, user-friendly, and fast.

Over the past several years the IEEE Geoscience and Remote Sensing Society, through its Data Archiving and Distribution Technical Committee, has sponsored several sessions at IGARSS, and the European Image Information Mining Coordination Group (IIMCG) has supervised four conferences jointly organized by the European Space Agency and the European Union Satellite Centre to provide a forum for IIM researchers. The focus of this special section is to describe some of the recent advances made in the field of IIM for Earth Observation in the development of tools, methods, and applications. It is composed of eight papers selected according to the standard review process of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. These contributions have been organized into four sections: Feature Extraction, Learning, Inference, and Semantic-Based Retrieval.

The section on *Feature Extraction* contains two manuscripts. The two contributions examine two different approaches for extracting a set of primitive features that can be used across a broad range of EO images and classes—artificial neural networks and wavelets. Image retrieval based on a semantic query is dependent on the primitive features used to describe the classes (e.g., water, agriculture, mountains) in an image and can become computationally intensive if the number of features is large and the archive being searched has many images. The paper by Del Frate *et al.* presents new results of designing neural networks, specifically a multilayer perceptron for the extraction of land cover features from large volumes of EO images. One of the main difficulties of information extraction from large image volumes is the variability of the spectral signatures. The selection of the training data is an important factor to obtain the needed generalization for the large data set. The generalization power comes from the similarity in shape of the spectral signatures, and thus, biases (e.g., illumination, atmospheric effects, etc.) become irrelevant for the classification. The developed method is demonstrated for relevant case

Digital Object Identifier 10.1109/TGRS.2007.894401

studies on urban areas for both medium-resolution and high-resolution sensors, Landsat and QuickBird, respectively. The short communications paper by Shah *et al.* proposes a systematic approach to selecting an appropriate wavelet decomposition level that is based on the energy (i.e., frequency content) of the image. Four classes (forest, water bodies, agricultural land, and fallow land) and three sets of imagery were used to validate the applicability of the technique. In all cases, it was shown that the results obtained with an adaptive selection were more robust than using a fixed decomposition level.

The section on *Learning* has two contributions of high relevance for mining the new generation of meter-resolution EO images. They refer to the need of nonimage information or external knowledge to help in the creation of a context for understanding or extracting the image information. While EO image resolution is increasing, the information content in its geometrical (spatial) context aspects becomes more and more important. The article by Gautama *et al.* introduces an image mining method based on spatial queries between images and GIS data. An error-tolerant graph matching technique is introduced and used to find correspondences between the detected image information and the road vector data, thus enabling spatial queries. The work of Fercatu and Boujemaa is aiming at the use of human knowledge, as interpretation context, to help the image retrieval process. Mainly for high-resolution EO images, due to their increased information details, the content interpretation depends on some hypothesis or prior knowledge of the user. Thus, the same part of an image can be assigned to different classes by different users. The proposed method is based on optimizing the transfer of information between the user and the system, and focuses on the criterion employed by the system for selecting the images presented to the user in a relevance feedback loop. The learning process is based on a support vector machine algorithm with a specific kernel design to obtain invariance to scale.

The *Inference* methods are represented by Gueguen and Datcu's paper, which deals with the problem of how to infer, in an unsupervised manner, the models of EO data, and thus, how to discover their information content. The paper presents a new information bottleneck algorithm used jointly with Bayesian methods for inferring and selecting the models explaining the observed data. The developed algorithms are demonstrated for data mining of spatio–temporal structures in high-resolution satellite image time series obtained by the SPOT sensors observing on a daily base a suburban-agriculture site for a period of more than eight months.

The *Semantic-Based Retrieval* section has three contributions. These papers provide examples of systems being developed to provide the user community with access to EO archives using semantic queries. The paper by Shyu *et al.* describes the Geospatial Information Retrieval and Indexing System (GeoIRIS) that is designed to enable scalable processing and retrieval of a large volume of data by automatically preprocessing and indexing satellite images. Its application area is the intelligence community. The paper covers the design of the system framework, as well as its approaches to feature extraction, database indexing, information ranking, semantic modeling, and queries. The contribution by Li and Bretschneider proposes a context-sensitive Bayesian network for semantic inference of segmented scenes. The segmented scene's semantic concepts are inferred in a multistage process based on their spectral and textural characteristics as well as the semantics of adjacent regions. A comparison was made with an approach that uses the feature space directly to infer the region's semantics. The result was the proposed Bayesian network approach outperformed the direct approach. The paper by Molinier *et al.* presents a search engine for detection, retrieval, and semantic annotation of man-made structures and for monitoring their changes from multi-temporal high-resolution EO images. Robust feature extraction algorithms are used to describe image tiles. The proposed features are adapted to the nature of high-resolution images. The image tile size is also carefully chosen. The learning mechanism is based on self-organizing maps to train the input and assigns a model vector to a visual label from the image tiles. The method is demonstrated using QuickBird images, and the results are evaluated quantitatively.

ACKNOWLEDGMENT

The Guest Editors would like to thank the IEEE Geoscience and Remote Sensing Society for its support to publishing this special section. Thanks are due to Professor J. A. Benediktsson, Editor-in-Chief, for his constant support and encouragement during the preparation of this special section's preparation. A special acknowledgement is also given to the reviewers of the papers, who played the most important role in the final selection and quality of the contributions published in this special section. Finally, we would like to thank Alison Larkin of the IEEE staff for her valuable assistance in all the phases of preparation of the special section.

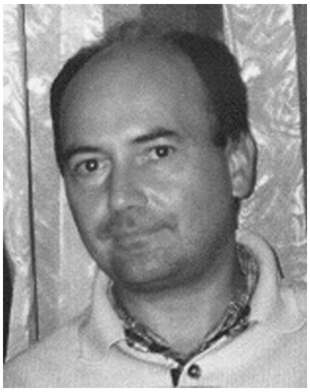
MIHAI DATCU, *Guest Editor*
Remote Sensing Technology Institute (IMF)
German Aerospace Center (DLR)
Oberpfaffenhofen, 82234 Wessling, Germany

Paris Institute of Technology
GET/Telecom Paris
75 013 Paris, France

SERGIO D'ELIA, *Guest Editor*
Service Support and Ground Segment Technology
European Space Agency
ESA Centre for Earth Observation (ESRIN)
Earth Observation Programme (EOP-GDR)
00044 Frascati, Italy

ROGER L. KING, *Guest Editor*
Bagley College of Engineering
Mississippi State University, Mississippi State
MS 39762-9544 USA

LORENZO BRUZZONE, *Associate Guest Editor*
Department of Information and Communication
Technology, University of Trento
38050 Trento, Italy



Mihai Datcu (SM'04) received the M.S. and Ph.D. degrees in electronics and telecommunications from the University "Politechnica" Bucharest UPB, Romania, in 1978 and 1986, respectively and the "Habilitation à diriger des recherches" from Université Louis Pasteur, Strasbourg, France, in 1999.

He holds a Professorship in electronics and telecommunications with the University "Politechnica" Bucharest UPB, since 1981. Since 1993, he has been a Scientist with the German Aerospace Center (DLR), Oberpfaffenhofen, Wessling, Germany. He is developing algorithms for model-based information retrieval from high-complexity signals and methods for scene understanding from Synthetic Aperture Radar (SAR) and interferometric SAR data. He is engaged in research related to information theoretical aspects and semantic representations in advanced communication systems. Currently, he is Senior Scientist and Image Analysis Research Group Leader with the Remote Sensing Technology Institute of DLR, Oberpfaffenhofen, the Coordinator of the CNES/DLR/ENST Competence Centre on Information Extraction and Image Understanding for Earth Observation, and Professor with the École Nationale Supérieure des Télécommunications Paris. His interests are in Bayesian inference, information and complexity theory, stochastic processes, model-based scene understanding image semantic coding, image information mining for applications in information retrieval and understanding of high-resolution SAR, and optical observations. He has held Visiting Professor appointments from 1991 to 1992 with the Department of Mathematics, of the University of Oviedo, Spain, and from 2000 to 2002, with the Université Louis Pasteur and the International Space University, both in Strasbourg, France. From 1992 to 2002, he had a longer Invited Professor assignment with the Swiss Federal Institute of Technology ETH, Zürich. In 1994, he was a Guest Scientist with the Swiss Center for Scientific Computing, Manno, and in 2003, he was a Visiting Professor with the University of Siegen, Germany. He is involved in advanced research programs for information extraction, data mining, and knowledge discovery and data understanding with the ESA, CNES, NASA, and in a variety of European projects.

Prof. Datcu is member of the European Image Information Mining Coordination Group.

Sergio D'Elia is currently with the European Space Agency (ESA), ESA Centre for Earth Observation (ESRIN), within the Earth Observation Programme (EOP-GDR), Head of the Service Support and Ground Segment Technology Section, Frascati, Italy. He started his career in November 1966 as an Electronic Hardware Designer and then became a Software Analyst. Since 1999, he has managed the Research and Technology Development Section, Earth Observation Programme, where he contributes to the definition of the European RTD requirements for the Earth observation ground segment and manages a team in charge of implementing the related plan (concerned areas include user interfaces, information systems, payload planning, information mining, feature extraction, archiving, processing, distribution, automation, service support, GRID, etc.). From 1975 (when he joined ESA) until 1986, he has served in the ESA's Information Retrieval Service, managing part of the software development of QUEST, the ESA's information retrieval system. From 1966 to 1975, he was employed in the research and development branches of various leading Italian research centers and industries (CNR, CSM, Contraves, and Selenia), with responsibilities that focused on electronic design of subsystems in support of various research topics, radar applications, or scientific satellite exploitation.

Roger L. King (M'73–SM'95) received the B.S. degree in electrical engineering from West Virginia University, Morgantown, in 1973, the M.S. degree in electrical engineering from the University of Pittsburgh, Pittsburgh, PA, in 1978, and the Ph.D. degree in engineering from the University of Wales, Cardiff, U.K., in 1988.

He began his career with Westinghouse Electric Corporation, but soon moved to the U.S. Bureau of Mines Pittsburgh Mining and Safety Research Center. In 1988, he accepted a position at the Department of Electrical and Computer Engineering, Mississippi State University, where he now holds the position of Giles Distinguished Professor. At the Mississippi State University, he presently serves as the Associate Dean for Research and Graduate Studies in the Bagley College of Engineering.

Dr. King is presently is a member of the IEEE GRSS AdCom. Over the last 30 years, he has served in a variety of leadership roles with the IEEE Industry Applications Society, Power Engineering Society, and Geosciences and Remote Sensing. He is a Registered Professional Engineer in the State of Mississippi. He has received numerous awards for his research including the Department of Interior's Meritorious Service Medal.

Lorenzo Bruzzone (S'95–M'98–SM'03) received the Laurea Specialistica (M.S.) degree (*summa cum laude*) in electronic engineering and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

From 1998 to 2000, he was a Postdoctoral Researcher with the University of Genoa. From 2000 to 2001, he was an Assistant Professor with the University of Trento, Trento, Italy, where he was an Associate Professor from 2001 to 2005. Since March 2005, he has been a Full Professor of telecommunications with the University of Trento, where he currently teaches remote sensing, pattern recognition, and electrical communications. He is currently the Head of the Remote Sensing Laboratory, Department of Information and Communication Technology, University of Trento. He is the author (or coauthor) of more than 150 scientific publications, including journals, book chapters, and conference proceedings. His current research interests are in the area of remote sensing image processing and recognition (analysis of multitemporal data, feature selection, classification, regression, data fusion, and machine learning). He conducts and supervises research on these topics within the frameworks of several national and international projects. Since 1999, he has been appointed Evaluator of project proposals for the European Commission. He is a Referee for many international journals and has served on the scientific committees of several international conferences.

Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (Seattle, July 1998). He received the Recognition of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Best Reviewers in 1999 and was a Guest Editor of a Special Issue of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING on the subject of the analysis of multitemporal remote sensing images (November 2003). He was the General Chair and Cochair of the First and Second IEEE International Workshop on the Analysis of Multi-temporal Remote Sensing Images. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is a member of the Scientific Committee of the India–Italy Center for Advanced Research. He is also a member of the International Association for Pattern Recognition and of the Italian Association for Remote Sensing (AIT).