4

Vision Goes Symbolic Without Loss of Information Within the Preattentive Vision Phase: The Need to Shift the Learning Paradigm from Machine-Learning (from Examples) to Machine-Teaching (by Rules) at the First Stage of a Two-Stage Hybrid Remote Sensing Image Understanding System, Part I: Introduction

Andrea Baraldi
Department of Geography, University of Maryland,
College Park, Maryland,
USA

1. Introduction

One traditional, although visionary goal of the remote sensing (RS) community is the development of operational satellite-based measurement systems suitable for automating the quantitative analysis of large-scale spaceborne multi-source multi-resolution image databases (Gutman et al., 2004). In past years this goal was almost exclusively dealt with by research programs focused on land cover (LC) and land cover change (LCC) detection at global scale (Gutman et al., 2004) (pp. 451, 452). In recent years the objective of developing operational satellite-based measurement systems has become increasingly urgent due to multiple drivers. While cost-free access to large-scale low spatial resolution (SR) (above 40 m) and medium SR (from 40 to 20 m) spaceborne image databases has become a reality (GEO, 2005; GEO, 2008a; GEO, 2008b; Gutman et al., 2004; Sart et al., 2001; Sjahputera et al., 2008), in parallel, the demand for high SR (between 20 and 5 m) and very high SR (VHR, below 5 m) commercial satellite imagery has continued to increase in terms of data quantity and quality, which has boosted the rapid growth of the commercial VHR satellite industry (Sjahputera et al., 2008). In this scientific and commercial context an increasing number of on-going international research projects aim at the development of operational services requiring harmonization and interoperability of Earth observation (EO) data and derived information products generated from a variety of spaceborne imaging sensors at all scales - global, regional and local. Among these on-going programs it is worth mentioning the Global EO System of Systems (GEOSS) conceived by the Group on Earth Observations (GEO) (GEO, 2005; GEO, 2008b), the Global Monitoring for the Environment and Security (GMES), which is an initiative led by the European Union (EU) in partnership with the

European Space Agency (ESA) (ESA, 2008; GMES, 2011), the National Aeronautics and Space Administration (NASA) Land Cover and Land Use Change (LCLUC) program (Gutman et al., 2004) (p. 3) and the U.S. Geological Survey (USGS)-NASA Web-Enabled Landsat Data (WELD) project (USGS & NASA, 2011).

Unfortunately, to date, the increasing rate of collection of EO imagery of enhanced spatial, spectral and temporal quality outpaces the automatic or semi-automatic capability of generating information from huge amounts of multi-source multi-resolution RS data sets (Gutman et al., 2004). This may explain why the percentage of data downloaded by stakeholders from the ESA EO image archives is estimated at about 10% or less (D'Elia, 2009).

If productivity in terms of quality, quantity and value of high-level output products generated from input EO imagery is low, this is tantamount to saying that existing scientific and commercial RS image understanding (classification) systems (RS-IUSs), such as (Definiens Imaging GmbH, 2004; Esch et al., 2008; Richter, 2006), score poorly in operational contexts (Tapsall et al., 2010). For example, RS-IUSs capable of proving their competitiveness at local/regional scale, such as the inductive supervised (labeled) data learning Support Vector Machines (SVMs) (Bruzzone & Carlin, 2006; Bruzzone & Persello, 2009), typically lack robustness and scalability for seamless application to LC and LCC problems at national, continental and global scale. As an example of these difficulties the interested reader may refer to (Chengquan Huang et al., 2008), where an SVM training algorithm and model selection strategies are applied to every image of a multi-temporal image mosaic at global scale. If the conjecture that existing RS-IUSs are affected by low productivity holds in general, it applies in particular to two-stage segment-based RS-IUSs which have recently gained widespread popularity and are currently considered the state-of-the-art in both scientific and commercial RS image mapping applications (Castilla et al., 2009; Mather, 1994). In literature the conceptual foundation of two-stage segment-based RS-IUSs is well known as geographic (2-D) object-based image analysis (GEOBIA), including a so-called iterative geographic OO image analysis (GEOOIA) approach (Baatz et al., 2008) (Hay & Castilla, 2006), also called object-oriented (image) analysis (OOA) (Castilla et al., 2008).

To summarize, in operational contexts (other than toy problems at small spatial scale and coarse semantic granularity) a RS-IUS can be considered as a low performer when at least one among several operational quality indicators (QIs) scores low. In (Baraldi et al., 2010a), a set of QIs eligible for use with an operational RS-IUS comprises the following: degree of automation (equivalent to ease of use; it is monotonically decreasing with the number of system-free parameters to be user-defined), classification and spatial accuracies (Baraldi et al., 2005), efficiency (e.g., computational time, memory occupation), robustness to changes in input parameters, robustness to changes in the input data set, scalability, timeliness (defined as the time span between data acquisition and high-level product delivery to the end user; it increases monotonically with manpower and computing time) and economy. In RS common practice, one or many of the aforementioned QIs of existing RS-IUSs tend to score low at local to global scale. This observation appears in line with a well-known opinion by Zamperoni according to which computer vision (CV) remains, to date, far more problematic than might be reasonably expected (Zamperoni, 1996). In

addition to CV, other scientific disciplines such as Artificial Intelligence (AI)/Machine Intelligence (MAI) and Cybernetics/Machine Learning (MAL), whose origins date back to the late 1950s, still remain unable to provide their ambitious cognitive objectives with operational solutions (Diamant, 2005; Diamant, 2008; Diamant, 2010a; Diamant, 2010b).[1]

To outperform existing scientific and commercial image understanding approaches, a new trend of research and development is found in both CV (Cootes and Taylor, 2004) and RS literature (Mather, 1994; Matsuyama & Shang-Shouq Hwang, 1990; Pekkarinen et al., 2009). This new trend aims at developing novel hybrid models for retrieving sub-symbolic (sensory, non-semantic, objective) continuous variables (e.g., leaf area index, LAI) and symbolic (categorical, semantic, subjective) discrete variables (e.g., land cover types) from optical multi-spectral (MS) imagery. By definition, hybrid models combine both statistical (inductive, bottom-up, fine-to-coarse, driven-without-knowledge, learning-from-examples) and physical (deductive, top-down, coarse-to-fine, prior knowledge-based, learning-by-rules) models to take advantage of the unique features of each and overcome their shortcomings (Matsuyama & Shang-Shouq Hwang, 1990; Shunlin Liang, 2004).

The original contribution of this work is to revise, integrate and enrich previous analyses found in related papers about recent developments in the design and implementation of an operational automatic multi-sensor multi-resolution near real-time two-stage hybrid stratified hierarchical RS-IUS (Baraldi et al., 2006a; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b). These novel developments encompass the four levels of analysis of an information processing system (Baraldi, 2011a; Marr, 1982), namely: (i) computational theory (system architecture), (ii) knowledge/information representation, (iii) algorithm design and (iv) implementation.

Starting from these recent achievements the present work provides an in-depth analysis of Emanuel Diamant's works including original speculations on the conceptual framework of MAI together with image segmentation and edge detection algorithms provided as proofs of his concepts (Diamant, 2005; Diamant, 2008; Diamant, 2010a; Diamant, 2010b). To overcome the conceptual and algorithmic drawbacks highlighted in Diamant's works, this manuscript proposes revised/new definitions of the following concepts: objective continuous sub-symbolic sensory data, continuous physical information, subjective discrete semi-symbolic data structure, discrete semantic-square (semantic^2) information and prior knowledge base. Continuous physical information is defined as a hierarchical description (multi-scale encoding/decoding or intra-scale transcoding) of an objective continuous sensory data set based on a given mathematical vocabulary/language, e.g., a fast Fourier transform (FFT) of a time signal. Discrete semantic^2 information is naturally (automatically, instantaneously) generated from the simultaneous combination of three components: (I) an objective continuous sensory data set, (II) an external subjective supervisor (observer) and (III) his/her own subjective prior ontology (model of the (3-D) world existing before looking at the objective sensory data at hand) whose hierarchical form is equivalent to that of a story in a natural language, comprising a title, an abstract, sections, paragraphs, sentences and words. In practical contexts these definitions imply the following.

---

[1] In Italian, acronym AI reminds of the English expression: 'ouch'. Acronym MAI means 'never'. Acronym MAL means 'pain'. Acronym MAT means 'fool'. These choices are arbitrary, but not by chance. Ancient Latins used to say: Nomen est omen... (meaning: 'true to its name').

a. It is impossible to *extract* semantic² information from objective continuous sensory data because the latter, *per se*, are provided with no semantics at all.
b. It is possible to *correlate* discrete semantic² information to objective continuous sensory data. Unfortunately, correlation between continuous sensory data and a finite and discrete set of categorical variables, corresponding to independent random variables generating separable data structures (data aggregations, data clusters, data objects), is low in real-world RS image mapping problems at large data scale or fine semantic granularity, other than toy problems at small data scale and coarse semantic granularity. This low correlation effect is due to the combination of two factors.

    i. According to the *central limit theorem* the distribution of the sample average of n independent and identically distributed (iid) random variables (corresponding to, say, categorical variables) approaches the normal distribution, featuring no "distinguishable" data sub-structure, as the sample size n increases. In other words, the separability of "distinguishable" data structures in a given measurement space of a given objective sensory data set is monotonically non-increasing (i.e., it decreases or remains equal) with the finite number of discrete semantic concepts (e.g., land cover classes) involved with the cognitive (classification) problem at hand.
    ii. In a given measurement space, within-class variability (vice versa, inter-class separability) is monotonically non-decreasing (i.e., it increases or remains equal) (vice versa, non-increasing) with the magnitude of the sample set per categorical variable when this variable-specific sample set size is "large" according to large-sample statistics (although large sample is a synonym for 'asymptotic' rather than a reference to an actual sample magnitude, a sample set cardinality of 30÷50 samples per random variable is typically considered sufficiently large that, according to a special case of the central limit theorem, the distribution of many sample statistics becomes approximately normal). For example, in (Chengquan Huang et al., 2008), where a time-consuming SVM training and classification model selection strategies are applied to every image of a world-wide RS image mosaic to separate forest from non-forest pixels, a so-called training data automation (TDA) procedure identifies a forest peak in a one-band first-order statistic (histogram) of a local image window. The size of this local image window must be fine-tuned based on heuristics because the inter-class spectral separability between classes forest and non-forest (vice versa, within-class variability) decreases (vice versa, increases) monotonically with the local window size above a certain (empirical) threshold (minimum window size, below which the collected sample is not statistically significant).

Some practical conclusions of potential interest to the RS, CV, AI and MAL communities stem from these speculations. Firstly, in operational contexts (e.g., RS image classification problems at national, continental and global scale), other than toy problems (e.g., RS image mapping at coarse spatial resolution and local/regional scale), inductive classifiers capable of learning from a finite labeled data set should be considered structurally inadequate to correlate (rather than extract, see this text above) discrete semantic² information with objective sensory data provided, *per se*, with no semantics at all.

Secondly, to increase the operational QIs of existing two-stage hybrid RS-IUSs, any first-stage inductive MAL-from-examples approach should be replaced by a deductive Machine Teaching (MAT)-by-rules sub-system capable of generating a preliminary classification first

stage in the Marr sense (Baraldi et al., 2006a; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b; Marr, 1982). As a proof of this concept the operational automatic prior knowledge-based multi-sensor multi-resolution near real-time Satellite Image Automatic Mapper™ (SIAM™) is selected from existing literature (Baraldi et al., 2006a; Baraldi et al., 2010a; Baraldi et al., 2010b; 1 Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b).

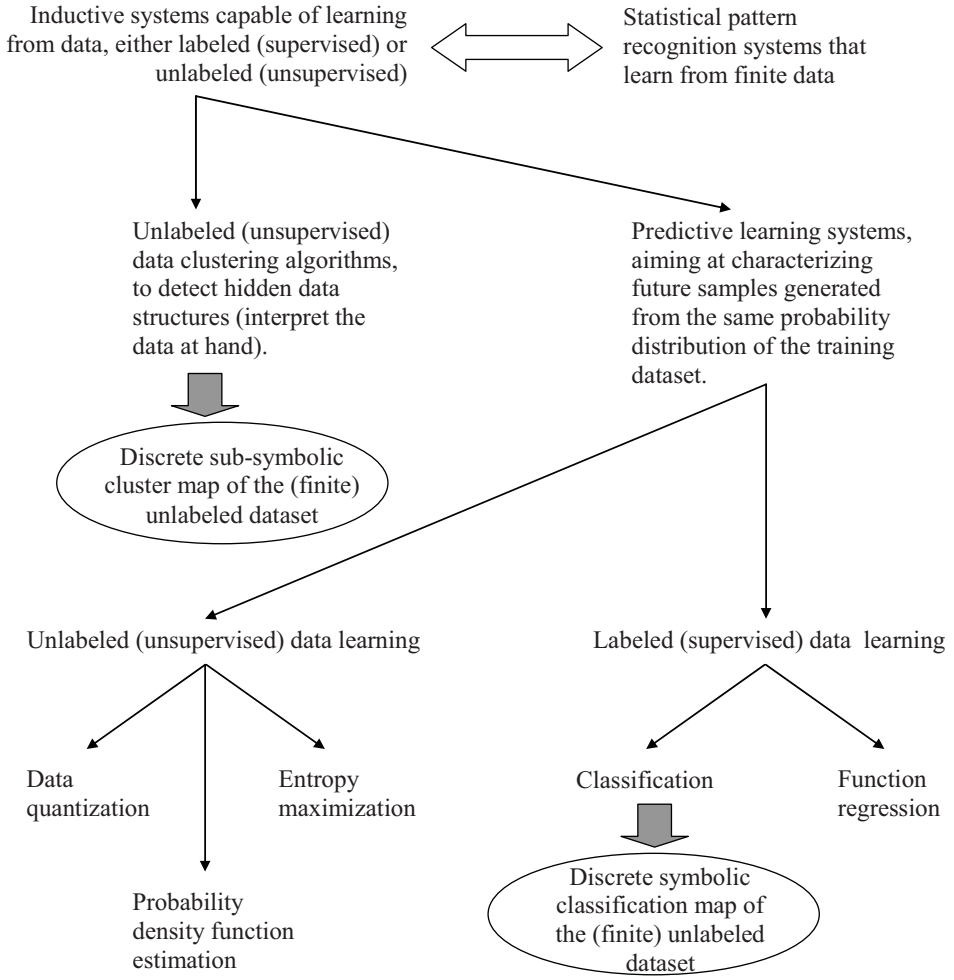

Fig. 1. The taxonomy of statistical pattern recognition systems proposed in (Baraldi et al., 2006b). Clustering algorithms and classification systems map an unlabeled input data sample into a discrete and finite set of sub-symbolic and symbolic labels, respectively. These discrete output maps are called (sub-symbolic) cluster maps (consisting of, say, cluster 1, cluster 2, etc.) and (symbolic) classification maps (consisting of, say, symbolic labels such as land cover classes broad-leaf forest, needle-leaf forest, etc.), respectively.

Thirdly, in RS-IUSs, MAL-from-data algorithms, either labeled (supervised) or unlabeled (unsupervised), either context-insensitive (e.g., pixel-based) or context-sensitive (e.g., 2-D object-based), should be adapted to work on a driven-by-knowledge stratified (semantic masked/layered) basis and moved to the second stage of a novel two-stage stratified hierarchical hybrid RS-IUS architecture recently proposed in RS literature (Baraldi et al., 2006a; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b).

The rest of this work is organized as follows. For publication reasons it consists of Part I and Part II. In Part I Section 2 related works, concepts and definitions are revised to provide this multi-disciplinary study with a significant survey value and make it self-contained. Part I Section 2 includes the following sub-sections: definitions and synonyms involved with inductive and deductive inference mechanisms (see Part I Section 2.1), a critical review of the history of AI/MAI and Cybernetics/MAL including a summary of Diamant's definitions of objective data, physical information, semantic information, knowledge and intelligence (refer to Part I Section 2.2), a definition of the cognitive process of vision (see Part I Section 2.3), a critical analysis of the inherent ill-posedness of inductive data learning algorithms (see Part I Section 2.4), a review of Diamant's image segmentation and contour detections algorithms presented as proofs of his concepts summarized in Part I Section 2.2 (refer to Part I Section 2.5), a discussion of the four levels of understanding of a RS-IUS (see Part I Section 2.6), a presentation (see Part I Section 2.7) of the Quality Assurance Framework for EO (QA4EO) guidelines (GEO/CEOSS, 2008) delivered by the Working Group on Calibration and Validation (WGCV) of the Committee of Earth Observations (CEOS), the space arm of the Group on Earth Observations (GEO) (GEO, 2005; GEO, 2008b), and a list of operational QIs of an RS-IUS (refer to Part I Section 2.8).

Part II includes a review session (see Part II Section 2) and an original contribution (from Part II Section 3 to Part II Section 7). In Part II Section 2 different families of existing RS-IUSs, namely, multi-agent hybrid RS-IUSs, two-stage segment-based RS-IUSs and two-stage stratified hierarchical hybrid RS-IUSs, are compared at the architectural level of analysis (refer to Part I Section 2.6). Part II Section 3 discusses theoretical inconsistencies and algorithmic drawbacks found in Diamant's works (discussed in Part I Section 2.2 and Part I Section 2.5, respectively). Revised/novel definitions of objective continuous sensory data, continuous physical information, discrete semantic^2 information and prior knowledge are provided in Part II Section 4. In Part II Section 5 practical consequences of the novel definitions provided in Part II Section 4 are considered for CV, AI and MAL applications. Part II Section 6 presents the operational automatic multi-sensor multi-resolution near real-time SIAM™ as a proof of the original concepts proposed in this work. Conclusions are reported in Part II Section 7.

2. Related works, concepts, definitions and synonyms

To provide this multi-disciplinary paper with a significant survey value and make it self-contained, a variety of related works, concepts and definitions collected from AI, MAL, CV and RS literature are revised in this section.

2.1 Inference mechanisms: Deductive top-down coarse-to-fine physical models and inductive bottom-up fine-to-coarse statistical models

Starting from classical philosophy to end up with MAL it is well known that the general notion of inference (learning) comprises two types of learning mechanisms.

1. "Induction, i.e., progressing from particular cases (e.g., training data) to general (e.g., estimated dependency or model)" (Cherkassky & Mulier, 2006). *Inductive inference systems* are also called *inference systems capable of learning-from-examples*, *bottom-up*, *fine-to-coarse*, *data-driven*, *driven-without-knowledge*, *statistical models, statistical pattern recognition systems* (Matsuyama & Shang-Shouq Hwang, 1990; Shunlin Liang, 2004). Statistical models are capable of learning from either *labeled* (*supervised*) or *unlabeled* (*unsupervised*) data, refer to Fig. 1. An inductive learning-from-data approach is not influenced by prior knowledge concerning the cognitive problem (e.g., our prior knowledge about an image), subjective desires (e.g., what information we are aiming to extract from an image) or subjective expectations (e.g., what we expect to see in an image). In the words of Cherkassky and Mulier, "induction amounts to forming generalizations from particular true facts. This is an inherently difficult (ill-posed) problem and its solution requires *a priori* knowledge in addition to data" (Cherkassky & Mulier, 2006) (p. 39). To summarize, inductive data learning problems are inherently ill-posed and require *a priori* knowledge in addition to (either labeled or unlabeled) sensory data to become better posed.

2. "Deduction, i.e., progressing from general (e.g., model) to particular cases (e.g., output values)" (Cherkassky & Mulier, 2006). *Deductive inference systems* are also called *inference systems capable of learning-by-rules*, *top-down*, *coarse-to-fine*, *model-driven*, *prior knowledge-based, driven-by-knowledge*, *physical models, physical pattern recognition systems* (Matsuyama & Shang-Shouq Hwang, 1990; Shunlin Liang, 2004), see Fig. 1. Physical models are abstracts of reality. They consist of prior knowledge of the physical laws of the (3-D) world which is available before (prior to) looking at the objective sensory data at hand.

As output, statistical and physical quantitative models of the (3-D) world (e.g., quantitative models of land surfaces observed from space) generate either *continuous sub-symbolic variables* (e.g., LAI) or *discrete symbolic (categorical) variables* (e.g., land cover types).

In addition to the synonyms presented above, the following terms are considered synonyms in the rest of this paper (Matsuyama & Shang-Shouq Hwang, 1990; Shunlin Liang, 2004).

- *Sub-symbolic, non-semantic, sensory, instantaneous, continuous, numerical, quantitative, objective, absolute, varying variables or sensations.*
- *Symbolic, discrete and semantic, categorical, linguistic, qualitative, subjective, abstract, vague, persistent, stable variables or percepts, concepts, classes of (3-D) objects in the (3-D) world, (3-D) object-models.*

In RS data applications, quantitative models are traditionally sorted into three major categories: *statistical*, *physical* and *hybrid*, whose main advantages and limitations are so well known in existing literature as to be summarized by Shunlin Liang in the following few words (Shunlin Liang, 2004).

a. Statistical models are inductive data learning systems (refer to this text above). Therefore, they are inherently difficult to solve (ill-posed) and their solution requires *a priori* knowledge in addition to data (Cherkassky & Mulier, 2006). Statistical pattern recognition systems are based on *correlation relationships* between objective sensory data (e.g., RS imagery) and either continuous (e.g., LAI) or categorical (e.g., land surface)

variables. Statistical models are easy to develop, e.g., a human expert is not required to search for an explicit deterministic function, if any, between, say, a target physical variable (e.g., LAI) and sensory data. However, they are effective for summarizing local data exclusively, i.e., they are usually (always?) site-specific (Shunlin Liang, 2004). For example, in RS common practice no machine capable of learning from either unlabeled or labeled data scores high in operational contexts such as satellite image mapping at national/ continental/ global scale. As a proof of this concept, in (Chengquan Huang et al., 2008), a time-consuming SVM (Bruzzone & Carlin, 2006) training and classification model selection strategies are enforced for every RS image in a world-wide image mosaic. In addition, supervised data learning algorithms, either context-insensitive (e.g., pixel-based) or context-sensitive (e.g., (2-D) object-based (Definiens Imaging GmbH, 2004; Esch et al., 2008)), require the collection of reference training samples which are typically scene-specific, expensive, tedious, difficult or impossible to collect (Gutman et al., 2004). This means that in practical RS data applications where supervised data learning algorithms are employed, the cost, timeliness, quality and availability of adequate reference (training/testing) datasets derived from field sites, existing maps and tabular data have turned out to be the most limiting factors on RS data product generation and validation (Gutman et al., 2004). Finally, since statistical models are inherently ill-posed, they are difficult to maintain, adapt, modify and scale according to changing input data sets, sensor specifications and/or user requirements. For example, the free parameter selection phase of any image segmentation algorithm tends to be difficult because: (i) it is based on heuristic (empirical) criteria (correlation relationships) and (ii) due to its inherent ill-posedness (artificial insufficiency (Matsuyama & Shang-Shouq Hwang, 1990)), any image segmentation algorithm is site-specific and simultaneously affected by both omission and commission segmentation errors within each image at hand (Burr & Morrone, 1992; Corcoran & Winstanley, 2007; Corcoran et al., 2010; Delves et al., 1992; Hay & Castilla, 2006; Matsuyama & Shang-Shouq Hwang, 1990; Petrou & Sevilla, 2006; Vecera & Farah, 1997).

b. Physical models consist of prior knowledge concerning the physical laws of the (3-D) world which is available before looking at the objective sensory data at hand. They follow the physical laws of the real (3-D) world to establish *cause-effect relationships*. They have to be learnt by a human expert based on intuition, expertise and evidence from data observation. Thus, unfortunately, it takes a long time for human experts to learn physical laws of the real (3-D) world and tune physical models (Mather, 1994; Shunlin Liang, 2004). On the other hand, physical models are more intuitive to debug, maintain and modify than statistical models. In other words, if the initial physical model does not perform well, then the system developer knows exactly where to improve it by incorporating the latest knowledge and information. For example, with a non-adaptive decision-tree classifier it is easy to find the node of the decision process in which a misclassification error occurs. In practice, a non-adaptive decision-tree classifier is well-posed (i.e., every data sample is assigned a semantic label according to a specific rule set), but subjective (i.e., different system developers may generate different non-adaptive decision-tree classifiers in the same application domain), refer to this text above.

c. Hybrid models combine both statistical and physical models to take advantage of the unique features of each and overcome their shortcomings (refer to the two previous paragraphs) (Matsuyama & Shang-Shouq Hwang, 1990; Shunlin Liang, 2004).

2.2 Brief history of AI/MAI and Cybernetics/MAL

In every ML textbook and in the world wide web it is easy to find historical information on the multiple rises and falls of expectations and achievements in scientific disciplines such as Cybernetics/MAL and AI/MAI related to the inductive and deductive inference paradigms respectively (refer to Part I Section 2.1).

2.2.1 1940s, 1950s and 1980s: Bottom-up inductive Cybernetics/MAL

In the 1940s and 1950s, a number of researchers, mostly located at Princeton University and the Ratio Club in England, started exploring the connection between neurology and information theory to develop electronic networks capable of exhibiting rudimentary intelligence conceived as self-organizing network properties. This new scientific discipline, called Cybernetics, investigates the capability of complex distributed processing systems, consisting of multiple processing elements (agents) dynamically interacting in multiple ways based on simple local rules, to display emergent macro behaviors and persistent network structures from an input data flow, i.e., local rules lead to global network properties. For example, data regularities detected by a self-organizing network of processing elements are equivalent to a compression of input information with which the distributed system can provide an abstract representation of the external environment.

The key features of complex network systems adaptive to data are that: (i) to understand how it works, a self-organizing network must be run (learning by doing), which is to say that learning, intended as self-organizing network capability, emerges without anyone needing to define what learning and intelligence are all about, (ii) the global behavior outlasts any of the network processing elements (persistence of the whole over time), (iii) it is the competition among processing elements and their (lateral) connections which leads to the emergence of specialized network (sub-)structures; without competition all processing units would behave alike and no specializations of the units would evolve (Fritzke, 1997; Lawley, 2003; Martinetz & Schulten, 1994).

By the late 1950s, in spite of the low technological development of electronic devices, electronic networks such as W. Grey Walter's turtles and the Johns Hopkins Beast were considered eligible for proving the cybernetic concepts. However, during the 1960s, symbolic AI approaches had achieved great success at simulating high-level thinking in small demonstration programs. So, by 1960 approaches based on cybernetics were abandoned or pushed into the background.

Next, by the 1980s progress in symbolic AI seemed to stall. Many researchers started believing that symbolic systems would never be able to imitate all the processes of human cognition, such as perception, learning and pattern recognition. Again, a number of researchers looked for a "sub-symbolic" distributed approach capable of solving specific AI sub-problems. The basic idea was: "Why trouble oneself trying to grasp the principles of intelligence? Let us give the machine the chance to find (in a bottom-up approach) the best way to mimic intelligence" (Diamant, 2010b). In the middle 1980s interest in "connectionism"

in general and so-called artificial neural networks in particular was revived by the works of David Rumelhart and others who focused on Multi-Layer Perceptrons (MLPs) and their Back-Propagation (BP) parameter adaptation algorithm. These and other distributed processing approaches, such as fuzzy learning systems and evolutionary computation, are now studied collectively by the emerging discipline of MAL (also called computational intelligence).

Finally, from the 1990s to date, MAL has achieved its greatest successes due to a combination of factors: the increasing computational power and memory capacity of computers, a greater emphasis on solving specific "tractable" MAL sub-problems and a new commitment by researchers to solid mathematical/statistical methods (Alpaydin, 2010; Bishop, 1995; Cherkassky & Mulier, 2006; Duda et al., 2001; Mitchell, 1997). In practice, once its first idealistic objective failed, MAL has been "broken into pieces, disintegrated and fragmented into many partial tasks and goals" to make its problem domain more "tractable" (Diamant, 2010b).

2.2.2 1956-1974, 1980s to date: Top-down deductive AI/MAI

Starting from the seminal work of Turing in 1950, the origin of AI dates back to the summer of 1956 when a conference on the campus of Dartmouth College was attended by John McCarthy, Marvin Minsky, Allen Newell and Herbert Simon who became the leaders of AI research for many decades. John McCarthy, who coined the term in 1956, defines AI as "the science and engineering of making intelligent machines" (Diamant, 2010b).

Intelligent agents must be able to set goals and achieve them by making choices that maximize the utility (or "value") of the available choices. To be termed intelligent these agents must be able to make predictions about how their actions will affect the present status of the world. This means they need a way to represent the current status of the world, to make predictions about the world's future status as a consequence of their actions, to have a periodical check to see if the world status matches their predictions and to change their plan as this becomes necessary, thus requiring the agent to reason under uncertainty.

Back in 1956 the excitement and hopes to reach AI goals in a short time were quite high. Herbert Simon predicted that "machines will be capable, within twenty years, of doing any work a man can do" (Diamant, 2010b). Marvin Minsky agreed by writing that "within a generation ... the problem of creating 'artificial intelligence' will substantially be solved". Reported by Diamant (Diamant, 2010b), Steve Grand sayed that “Rodney Brooks has a copy of a memo from Marvin Minsky in which he suggested charging an undergraduate for a summer project with the task of solving vision. I don’t know where that undergraduate is now, but I guess he hasn’t finished yet”.

Many of the cognitive problems AI was expected to solve require extensive prior knowledge of the (3-D) world. A representation of "what exists in the (3-D) world" pertaining to the cognitive problem at hand is called *world model* (Matsuyama & Shang-Shouq Hwang, 1990) or *ontology* (borrowing a word from traditional philosophy). The graphical representation and implementation of an ontology is twofold.

- An *inverted tree* whose leaves are at the bottom level (layer 0), where semantic primitives (hereafter called *semi-concepts*) are found (Diamant, 2005; Diamant, 2010a; Diamant, 2010b; Diamant, 2008).

- A *semantic net* (*concept net*) is defined as a graph, either directed or non-oriented, either cyclic or acyclic, consisting of nodes linked by edges. Nodes represent concepts, i.e., classes of (3-D) objects in the world (see Part I Section 2.1), while edges represent relations, e.g., PART-OF, A-KIND-OF, spatial relations either topological (e.g., adjacency, inclusion) or non-topological (e.g., distance, angle), temporal transitions between nodes, physical model-based relationships between causes and effects, etc. (Hudelot et al., 2008; Matsuyama & Shang-Shouq Hwang, 1990; Pakzad et al., 1999).

Unfortunately, the number of atomic facts about the world that an average person knows is astronomical. It means that AI projects whose goal is to build a complete knowledge base of commonsense knowledge would require enormous amounts of laborious ontological engineering where one abstract concept must be built, by hand, at a time. In practice, it takes a long time for human experts to define ontologies, learn physical laws of the real (3-D) world and tune physical models based on human intuition, domain expertise and evidence from data observation. Within a decade or so it became clear that AI problems were immense, maybe even intractable. In 1974, in response to ongoing criticism and pressure to fund more productive projects, the U.S. and British governments cut off all exploratory research related to AI.

However, in the 1970s, computers with large memories became available. This drove AI researchers to began building prior knowledge into AI problem-specific "tractable" applications. In the early 1980s this led to the first commercial success of expert systems, a form of AI programs that simulated the knowledge base and analytical skills of human experts. By 1985 the market for AI reached over a billion dollars. At the same time, Japan's fifth generation computer project inspired the U.S and British governments to restore funding for academic research in the AI field. However, beginning with the collapse of the Lisp Machine market in 1987, AI once again fell into disrepute and a second, longer lasting, AI winter began.

Finally, from the 1990s to date, AI achieved its greatest successes, albeit somewhat behind the scenes. This success was due to a combination of factors, which are not surprisingly the same as those working in favor of the recent achievements of MAL (also refer to Part I Section 2.2.1), namely: the increasing computational power and memory capacity of computers, a greater emphasis on solving specific "tractable" AI sub-problems, a new commitment by researchers to solid mathematical/statistical methods and more rigorous scientific standards (Alpaydin, 2010; Bishop, 1995; Cherkassky & Mulier, 2006; Duda et al., 2001; Mitchell, 1997), and the creation of new ties between AI and other fields working on similar problems, such as MAL, knowledge representation (e.g., fuzzy logic) and uncertainty engineering (e.g., sensitivity analysis, error propagation). For example, a major goal of contemporary AI is to have the computer understand enough concepts to be able to learn by reading from sources like the internet, and thus be able to add to its own ontology. This is called Natural Language Processing, which gives machines the ability to read and understand the languages that humans speak.

Among the longest-standing AI questions that have remained unanswered, consider the following.

- Should AI simulate natural intelligence by studying psychology or neurology? Or is human biology as irrelevant to AI research as bird biology is to aeronautical engineering?

- In the attempt to develop hybrid inference systems where both statistical and physical models are combined to overcome their shortcomings (see Section 2.1), how, when and where do continuous sensory objective sub-symbolic data become discrete symbolic subjective information? This is the well-known *information gap* existing between (sub-symbolic, sensory, instantaneous, numerical, quantitative, absolute, non-semantic) sensations and (symbolic, linguistic, qualitative, vague, discrete and semantic, persistent, stable) percepts (refer to Part I Section 2.1), which has been thoroughly investigated in both philosophy and psychophysical studies of perception (Matsuyama & Shang-Shouq Hwang, 1990). In practice, "we are always seeing objects we have never seen before at the sensation level, while we perceive familiar objects everywhere at the perception level" (Matsuyama & Shang-Shouq Hwang, 1990).

2.2.3 Fundamental flaws responsible for AI and MAL derailment: The Diamant perspective

When did AI and MAL derail from their original and ambitious goals? Diamant's answer is: They did it right at their origin dating back to the late 1950s (refer to Part I Section 2.2.1 and Part I Section 2.2.2, respectively) due to the following fundamental flows (Diamant, 2010b).

a. The lack of proper definitions to distinguish between objective data, physical information, semantic information, knowledge and intelligence. These definitions deal with the well-known *information gap* between physical and semantic information thoroughly investigated in both philosophy and psychophysical studies of perception (see Part I Section 2.2.2). In Diamant's words: "In my view, philosophy is not a swear-word. Philosophy is a keen attempt to approach the problem from a more general standpoint, to see the problem from a wider perspective, and to yield, in such a way, a better comprehension of the problem's specificity and its interaction with other world realities. Otherwise we are ... prone to dead-ends and local traps" (Diamant, 2010b).
b. Misunderstanding of the very nature of semantic information. Unlike physical information, semantics is not a property of the raw data, but the property of an external observer who observes and scrutinizes the data. Since semantics is assigned to physical data structures by an external observer, it cannot be learned from the sensory data.

The Diamant explanations of these concepts are quoted below (Diamant, 2005; Diamant, 2008; Diamant, 2010a; Diamant, 2010b).

2.2.3.1 Kolmogorov's complexity theory

Among definitions of "data", "information", and "knowledge", the definition of information is the most controversial. To provide it, Diamant relies on Kolmogorov's complexity theory (actually developed independently by Kolmogorov, Chaitin, and Solomonoff), whose concern is: What is the best way to represent a single data object? What are the laws of minimizing the length of a description of a single data object? Such a short-length compressed description is the information that we are seeking about a particular data object.

Theoretically two extreme cases can be distinguished: (1) the elements of a data set are absolutely random and (2) the elements of a data set form "observable" data structures. In the first case the data set can be represented only by the original sequence of its data elements. In the second case the presence of observable data structures consisting of data elements can be taken into account, which leads to a more compact and concise

(compressed) description. In terms of Kolmogorov's theory, this compressed description (encoding) must be a trustworthy (which does not mean lossless) abstract (summary) of the original data set such that: (i) the abstract description length is definitely shorter than the original uncompressed data set description, (ii) the abstract description is sufficient to reconstruct (reproduce, re-establish, decode) the salient properties or regularities or distinguishable data structures or data objects in the original data set.

Kolmogorov's theory prescribes the way in which a data set description has to be created: Firstly, the most simplified and generalized data structures must be described. (Recall the Occam's Razor principle: Among all hypotheses consistent with the observation, choose the simplest one that is coherent with the data (Mitchell, 1997)). Then, as the level of generalization (vice versa, granularity) is gradually decreased (vice versa, becomes finer), more and more fine-grained data details (structures) can be revealed and described.

2.2.3.2 Diamant's definitions of objective data, physical information, semantic information, knowledge and intelligence

Diamant reviews two survey papers (Legg & Hutter, 2007; Zins, 2007), published in the year 2007, where definitions of data, information, knowledge and MAI are collected from existing literature for comparison purposes. In (Zins, 2007), 130 definitions of data, information and knowledge are provided by 45 scholars. In (Legg & Hutter, 2007), more than 70 definitions of MAI are collected. According to Diamant, "what these two collections undoubtedly exhibit... is that definitions offered by the leading scholars in each field have nothing in common among them, and therefore are of little use when it comes to our practical problem-solving" (Diamant, 2010b). As a result, Diamant is forced to search for his own definitions.

Starting from the Kolmogorov complexity theory (see Section 2.2.3.1), Diamant provides the following definitions about data, information and knowledge.

1. (Objective) "data is an agglomeration of elementary facts" (Diamant, 2010a).
2. (Physical and semantic) "information is a description" (based on a) "language and/or alphabet" (Diamant, 2010a).
3. (Physical and semantic) "information is a hierarchy of decreasing level descriptions" (Diamant, 2010a).
4. (Physical?) "information elicitation (extraction) does not require incorporation of any high-level knowledge" (Diamant, 2010b; Diamant, 2008).
5. "Two kinds of information must be distinguished: objective (physical) information and subjective (semantic) information.
    a. By physical information we mean the description of data structures that are discernable in a data set" (Diamant, 2010b). (Noteworthy,) "successful recovery and description of image structures (e.g., successful image segmentation) does not lead to image understanding. The (data) structures that are observed in an image reflect aggregations of nearby data elements on the basis of similarity among their physical attributes (e.g., color or brightness in visual signals, frequency and intensity in audio signals). These (are called) 'primary (data) structures' or 'physical (data) structures'" (Diamant, 2010a). "Physical information, being a natural property of the data, can be extracted instantly from the data and no special rule is needed for such a task accomplishment" (Diamant, 2010b). (It is) "physical information... the only information present in an image, and therefore the only

information that can be extracted from an image " (Diamant, 2008). (In other words,) "defining (primary data structures) is certainly a well-grounded procedure that does not raise any objections, because objective (physical) nature laws underpin such a procedure" (Diamant, 2010a) (refer to point 4. above).

To summarize, according to Diamant, *physical information*, non-semantic *primary data structures* and discernable non-semantic *image segments* are synonyms.

b. "By semantic information we mean the description of the relationships that may exist between the physical (data) structures of a given data set" (Diamant, 2010b). (In other words,) "'primary (data) structures'... undergo a further grouping and aggregation, which leads to formation of 'secondary (data) structures' (consisting of primary data structures) that can be called... 'semantic (data) structures'" (Diamant, 2010a)."Unlike physical information, semantics is not a property of the raw data. Semantics is assigned to physical data structures by an external observer who watches and scrutinizes the data... Semantics is a shared convention, a mutual agreement between the members of a particular group of viewers or users. Its assignment (to the primary data structures) has to be made on the basis of a consensus knowledge that is shared among the group members, and which an artificial semantic-processing system has to possess at its disposal... Therefore semantics cannot be learned straightforwardly from the raw data" (Diamant, 2010b). (In other words,) "the knowledge about the rules that underpin secondary (data) structures formation is a property of human observers and not an inherent property of the data" (Diamant, 2010a). (Since) "semantic information is a convention, an agreement, a property shared between a company of particular observers, it cannot be learned (from physical data) by any means. It can be exchanged, transferred, relocated between the group members, or between humans and intelligent machines (robots) collaborating with them in a working group, but it cannot be learned (from data)" (Diamant, 2010b). (This implies that) "MAL techniques are ... not applicable for the purposes of semantic information extraction (from the raw data set)... (Acquisition) of this knowledge presumes availability of a different and usually overlooked special learning technique, which would be best defined as Machine Teaching (MAT) – a technique that would facilitate externally-prepared-knowledge transfer to the system's disposal" (Diamant, 2010b).

To summarize, according to Diamant *semantic information* and semantic *secondary data structures*, generated from subjective aggregation (semantic labeling) of non-semantic primary data structures, e.g., *image segments*, by an external observer, are synonyms. In addition, what Diamant calls MAT is known in traditional AI as knowledge engineering, which is a process of codifying human knowledge into an expert system (Laurini and Thompson, 1992).

6. "Both physical and semantic information descriptions are similar in that: (1) they are character strings, (2) they are top-down coarse-to-fine hierarchies, and (3) they are implemented according to a certain vocabulary/language. There is only a small difference – physical information can be described in a variety of languages while semantic information can be represented only in a human natural language... Therefore the most suitable form of semantic information representation should be a narrative, a

story, a tale. The usual top-down hierarchical structure of such a story (a narrative, a tale) is well known from other linguistic studies. Moving top-down, a story comprises a story title, abstract, chapter or section partition, paragraph subdivision, separate phrases and sentences which end up with single words (congregations of letters) actually composing a phrase. Further structural descent leads in linguistics to syntaxes. But in our case – the lowest level of a semantic structure is stuffed with physical information which represents the physical structure of a meaningful object designated by the word in a phrase... At the lowest level of a semantic description (hierarchy) a physical information sub-hierarchy is always present" (Diamant, 2010a).

To summarize, according to Diamant *semantic information* comprises *physical information* at the lowest level of a semantic description (hierarchy) equivalent to an inverted tree (see Part I Section 2.2.2).

7. (Prior) "knowledge is memorized (semantic) information (stored in the system's memory, which incorporates physical information)" (Diamant, 2010b).
8. "Data is not information, but knowledge is information (semantic information memorized in system's memory)" (Diamant, 2010b).
9. "Intelligence (cognition) is the system's ability to process (semantic) information" (Diamant, 2010b).

Together with the aforementioned theoretical considerations, Diamant presents an unlabeled (unsupervised) multi-scale image segmentation algorithm and a single-scale unlabeled (unsupervised) image contour detector as proofs of his concepts (Diamant, 2005). A critical analysis of these theoretical and algorithmic contributions by Diamant can be found in Part II Section 3.

2.3 Vision as an ill-posed image understanding problem

The main role of a biological or artificial visual system is to backproject the information in the (2-D) image domain to that in the (3-D) scene domain (Matsuyama & Shang-Shouq Hwang, 1990). In greater detail, the goal of a visual system is to provide plausible (multiple) symbolic description(s) of the scene depicted in an image by finding associations between sub-symbolic (non-semantic, sensory, instantaneous, numerical, absolute, quantitative, varying, objective, see Part I Section 2.1) (2-D) image features or sensations with symbolic (semantic, subjective, linguistic, qualitative, vague, abstract, persistent, stable, see Part I Section 2.1) (3-D) objects (concepts or percepts) in the scene (e.g., a building, a road, etc.). Sub-symbolic (2-D) image features are either points or regions or, vice versa, region boundaries, i.e., edges, provided with no semantic meaning. In literature, (2-D) image regions are also called *segments*, *(2-D) objects*, *patches*, *parcels*, or *blobs* (Carson et al., 1997; Lindeberg, 1993; Yang & Wang, 2007).

There is a well-known *information gap* between symbolic information in the (3-D) scene and sub-symbolic information in the (2-D) image, e.g., due to dimensionality reduction and occlusion phenomena, see Fig. 2 (also refer to Part I Section 2.2.2 and Part I Section 2.2.3). This is called the *intrinsic insufficiency* of image features (Matsuyama & Shang-Shouq Hwang, 1990). This information gap is also related to the inherent ill-posedness of inductive inference (see Part I Section 2.1). It means that the problem of image understanding is inherently ill-posed and, consequently, very difficult to solve (Matsuyama & Shang-Shouq Hwang, 1990; Cherkassky & Mulier, 2006).

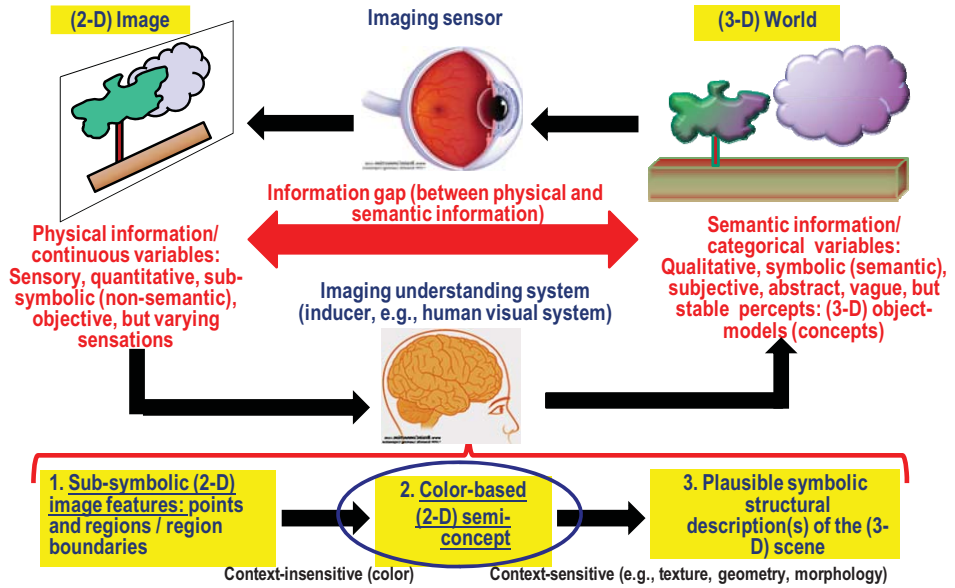

Fig. 2. Inherently ill-posed image understanding problem (vision). There is a well-known information gap between physical information and semantic information. This is the same information gap existing between (sub-symbolic, sensory, instantaneous, numerical, quantitative, absolute, non-semantic) sensations and (symbolic, linguistic, qualitative, vague, discrete and semantic, persistent, stable) percepts (concepts) which has been thoroughly investigated in both philosophy and psychophysical studies of perception. In practice, "we are always seeing objects we have never seen before at the sensation level, while we perceive familiar objects everywhere at the perception level" (Matsuyama & Shang-Shouq Hwang, 1990). The original automatic SIAM™ software button (executable), adopted as preliminary classification first stage of a novel two-stage stratified hierarchical hybrid RS-IUS architecture (see Part II, Section 2), generates as output a mutually exclusive and totally exhaustive set of symbolic spectral-based semi-concepts, also called spectral categories or land cover class sets, e.g., 'vegetation' (Baraldi et al., 2006a; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b). The semantic meaning of a spectral-based semi-concept is: (a) superior to zero, which is the semantic value of traditional sub-symbolic image features, namely, pixels, (2-D) image segments or edges, and (b) equal or inferior to the semantic meaning of target (3-D) land cover classes (e.g., needle-leaf forest), also called concepts or (3-D) object-models in the (3-D) world.

The aforementioned information gap coincides with the well-known *information gap* existing between (sub-symbolic, sensory, quantitative, objective, varying) sensations and (symbolic, semantic, qualitative, subjective, stable) percepts, traditionally investigated in both philosophy and psychophysical studies of perception (Matsuyama & Shang-Shouq Hwang, 1990) (see Part I Section 2.2.2).

In functional terms, biological vision combines preattentive (low-level) visual perception with an attentive (high-level) vision mechanism (Gouras, 1991; Kandel, 1991; Mason & Kandel, 1991).

1. Preattentive (low-level) vision extracts picture primitives based on general-purpose image processing criteria independent of the scene under analysis. It acts in parallel on the entire image as a rapid (< 50 ms) scanning system to detect variations in simple visual properties. It is known that the human visual system employs at least four spatial scales of analysis (Wilson & Bergen, 1979). Marr calls the output of the low-level vision first stage *primal sketch* or *preliminary map* (Marr, 1982).
2. Attentive (high-level) vision operates as a careful scanning system employing a focus of attention mechanism. Scene subsets, corresponding to a narrow aperture of attention, are looked at in sequence and each step is examined quickly (20–80 ms).

Finally, it is worth mentioning that, according to Marr, "vision goes symbolic almost immediately, right at the level of zero-crossing (primal sketch)... without loss of information" (Marr, 1982) (p. 343)**. In practice, Marr suggests the following.**

a. The output of preattentive vision (primal sketch) is symbolic. This is tantamount to saying that:
   - **vision goes symbolic within the preattentive vision phase,**
   - **the primal sketch is a preliminary semantic map whose symbolic labels belong to a finite and discrete set of 3-D object-classes or concepts in the real (3-D) world.**

b. The symbolic output of preattentive vision (refer to point (a) above) is lossless, i.e., when the input image is reconstructed from its semantic description, then small, but genuine image details (high spatial frequency image components) must be well preserved.

It is also noteworthy that, in contradiction with his own intuition about what functional properties characterize a biological vision system, the CV system implemented by Marr is unable to accomplish either of the two aforementioned goals (a) and (b). For example, the Marr pre-attentive vision module consists of a contour detector (zero-crossing) whose output is a sub-symbolic primal sketch. This is not at all surprising. It accounts in general for the customary distinction between a model and the algorithm used to identify it (Baraldi et al., 2010a; Baraldi, 2011a) (also refer to Part I Section 2.6) and, in particular, for the seminal nature of the conceptual work by Marr followed by his early dramatic death.

2.4 A few comments about the inherent ill-posedness of inductive MAL from either labeled or unlabeled data

Inductive machine learning from either labeled or unlabeled data (see Fig. 1) has been central to MAL research from the beginning. In particular, "induction amounts to forming generalizations from particular true facts. This is an inherently difficult (ill-posed) problem and its solution requires a priori knowledge in addition to data" (Cherkassky & Mulier, 2006) (p. 39), to make the ill-posed inductive learning-from-data problem better posed (see Part I Section 2.1). Unfortunately, although acknowledged by a significant portion of existing literature, the inherent ill-posedness of inductive MAL from either labeled or unlabeled data appears ignored or neglected by the majority of scientists and practitioners involved with MAL common practice.

2.4.1 Inherently ill-posed unlabeled data learning

Unlabeled (unsupervised) data learning is the ability to find discrete patterns or sub-symbolic labeled data structures in an input stream of unlabeled data vectors. Well-known

examples of discrete sub-symbolic data structures distinguishable in a stream of unlabeled data vectors are: (a) discrete sub-symbolic clusters (e.g., cluster 1, cluster 2, etc.) in a finite unlabeled data set belonging to a multi-dimensional measurement space and (b) discrete sub-symbolic (2-D) image segments (e.g., segment 1, segment 2, etc) found in a 2-D one-band (e.g., panchromatic) or multi-band (chromatic) image domain (see Fig. 1).

Inherently ill-posed unlabeled data clustering and image segmentation are further discussed below.

2.4.1.1 Inherently ill-posed unlabeled data clustering

Since the goal of clustering is to group the data at hand rather than to provide an accurate characterization of unobserved (future) samples generated from the same probability distribution, then the task of clustering may fall outside the framework of predictive learning (Cherkassky & Mulier, 2006). In spite of this, clustering analysis often employs unsupervised data learning approaches originally developed for vector quantization (such as the well-known k-means unsupervised data learning algorithm belonging to the family of the crisp competitive minimum-distance-to-means algorithms (Baraldi & Blonda, 1999a; Baraldi & Blonda, 1999b)), which is a predictive learning problem, see Fig. 1 (Cherkassky & Mulier, 2006).

Unlabeled data clustering is an inherently ill-posed data mapping problem. In fact, the goal of clustering is to separate a finite unlabeled dataset at hand into a finite and discrete set of "natural", hidden data structures on the basis of an often subjectively chosen measure of similarity/dissimilarity, i.e., a similarity measure chosen subjectively based on its ability to create "interesting" clusters (Backer & Jain, 1981; Baraldi & Alpaydin, 2002a; Baraldi & Alpaydin, 2002b; Cherkassky & Mulier, 2006; Fritzke, 1997). Thus, the subjective (ill-posed) nature of the nonpredictive data clustering problem precludes an absolute judgment as to the relative effectiveness of all clustering techniques (Backer & Jain, 1981). In spite of this, the inherent ill-posedness of unlabeled data clustering problems is not clearly stated in existing literature where, as a consequence, dozens of papers proposing alternative clustering algorithms are published every year (perhaps in search of a "final" best clustering algorithm which cannot exist…) (Xu & Wunsch II, 2005).

Crisp (hard) competitive minimum-distance-to-means algorithms, such as the k-means data quantization approach, try to minimize a sum-of-squares error function (Cherkassky & Mulier, 2006; Bishop, 1995). To reduce the risk of being trapped in a local minimum of the error function, soft-to-hard rather than hard competitive clustering algorithms have been conceived (Baraldi & Blonda, 1999a; Baraldi & Blonda, 1999b). In addition, it is well known that both crisp and fuzzy k-means data clustering algorithms cannot perform well with non-convex types of data, i.e., they are effective if and only if data clusters are hyperspherical (Duda et al., 2001). To overcome this problem, a k-means unsupervised data learning algorithm capable of defining automatically the number of clusters splits a non-convex data cluster, say, a data cluster shaped like a banana, into several hyperspheres. Thus, these hyperspheres should be linked to map the banana-like data cluster. To perform non-convex unlabeled data mapping, topologically preserving data clustering algorithms have been developed (Baraldi & Alpaydin, 2002a; Baraldi & Alpaydin, 2002b; Fritzke, 1997; Martinetz & Schulten, 1994).

In terms of degree of automation, which decreases monotonically with the number of system-free parameters to be user-defined, it is noteworthy that, to make the inherently ill-posed unsupervised data clustering problem better posed, every unsupervised data clustering algorithm requires at least one free parameter to be user-defined or fixed by the application developer based on heuristics. For example, it appears paradoxical that the well-known k-means vector quantizer, typically employed for unlabeled data clustering (refer to previous paragraphs), requires the user to pre-define the unknown number of unlabeled data clusters to be found in the finite unlabeled data set at hand.

In terms of computation time, unlabeled data clustering (either batch or on-line learning) is iterative (sub-optimal) in nature, therefore it is time-consuming with respect to prior knowledge-based one-pass data mapping algorithms (e.g., pattern-matching techniques).

In terms of effectiveness and robustness to changes in the input dataset, on-line (stochastic, sequential) learning unlabeled data clustering algorithms are typically subjected to local minima, e.g., they are sensitive to the order of presentation of the input data sequence. To enhance their robustness to changes in the order of presentation of the input sequence, semi-batch unlabeled data clustering algorithms have been developed (Wilson & Martinez, 2000).

2.4.1.2 Inherently ill-posed (2-D) image region extraction/contour detection

In literature, a so-called Low-Level Vision Expert (LLVE) (Matsuyama & Shang-Shouq Hwang, 1990) includes a battery of low-level sub-symbolic (non-semantic) general-purpose domain-independent inductive-learning (fine-to-coarse, bottom-up, driven-without-knowledge, see Part I Section 2.1) inherently ill-posed image processing (unlabeled data-driven) algorithms working at the signal level. This set of low-level image processing algorithms may comprise (Matsuyama & Shang-Shouq Hwang, 1990): edge-preserving noise filtering (Acton & Landis, 1997; Perona & Malik, 1990), either intensity- or color-based region/edge detection (Baraldi & Parmiggiani, 1996a; Canny, 1986), texture-based region/edge detection (Jain & Healey, 1998), region growing (Baraldi & Parmiggiani, 1996b), region extraction from not-close contours (Baraldi & Parmiggiani, 1995), etc.

In a (2-D) image domain, **region extraction is the dual problem of edge detection** and they are both inherently ill-posed visual tasks. In the rest of this paper, for simplicity's sake, in line with (Matsuyama & Shang-Shouq Hwang, 1990), all the aforementioned image processing operators are called "**segmentation**" algorithms. As output, an image segmentation algorithm generates *image features*, namely *points* and *regions* (also called segments, [2-D] objects, parcel or blobs (Carson et al., 1997; Lindeberg, 1993; Yang & Wang, 2007), also refer to Part I Section 2.3) or, vice versa, *region boundaries*, i.e., *edges*, provided with no semantic meaning. In general, a sub-symbolic image segment is: (1) made of connected pixels considered homogeneous in color and/or texture based on: (i) a subjective measure of similarity/dissimilarity and (ii) a subjective decision rule (e.g., thresholding), and (2) provided with a non-semantic label equivalent to a numerical segment-based identifier (integer value).

The inherent ill-posedness of any image segmentation algorithm is due to both systematic and accidental errors. The so-called *intrinsic insufficiency* of image segments is due to occlusion problems and dimensionality reduction (Matsuyama & Shang-Shouq Hwang, 1990) (refer to Part I Section 2.3). In addition, image segments are always affected by a so-

called *artificial insufficiency* (Matsuyama & Shang-Shouq Hwang, 1990) due to the image segmentation algorithm at hand. This latter source of segmentation errors is related to the well-known ***uncertainty principle*** **according to which, for any contextual (neighborhood) property, we cannot simultaneously measure that property while obtaining accurate localization** (Corcoran & Winstanley, 2007; Petrou & Sevilla, 2006).

In practical contexts the inherent ill-posedness of any image segmentation algorithm implies the following.

(a) In real-world image segmentation problems (other than toy problems), it is inevitable for erroneous segments to be detected while genuine segments are omitted (Matsuyama & Shang-Shouq Hwang, 1990) (p. 18).
(b) Any image segmentation algorithm must rely on user-defined segmentation-free parameters based on subjective (heuristic, empirical) criteria on a site-specific basis (see Part I Section 2.1). As a consequence, any image segmentation algorithm can be considered difficult to use, i.e., its degree of automation is low, while its robustness to changes in the input data set and changes in input parameters are both low.

To overcome these shortcomings many researchers in the field of cognitive psychology believe that object segmentation cannot be achieved in a completely bottom-up manner, which is tantamount to saying that segmentation and classification are strongly coupled (Corcoran & Winstanley, 2007; Corcoran et al., 2010; Vecera & Farah, 1997). In particular, Vecera and Farah proved that the process of human visual segmentation can be strongly influenced by top-down human (subjective) factors such as prior knowledge of the image at hand in addition to desires and expectations of an external observer (Vecera & Farah, 1997).

To date, the inherent ill-posedness of any image region/boundary detection algorithm is acknowledged by a relevant portion of the CV and RS communities (Burr & Morrone, 1992; Corcoran & Winstanley, 2007; Corcoran et al., 2010; Delves et al., 1992; Hay & Castilla, 2006; Matsuyama & Shang-Shouq Hwang, 1990; Petrou & Sevilla, 2006; Vecera & Farah, 1997). For example, Castilla *et al*. observe that (Castilla et al**.**, 2008): " Image understanding is a complex cognitive process for which we may still lack key concepts. In particular, most image segmentation methods have been developed heuristically without a deeper examination of the semantic implications of the segmentation process." Well-known image segmentation algorithms, including eCognition® by Definiens AG (Definiens Imaging GmbH, 2004), "... are conceptually inconsistent with the object-oriented approach (OOA)... an underlying hypothesis of any segmentation method is that there is a correspondence between radiometric similarity in the image and semantic similarity in the imaged landscape. Thus, it is expected that image objects (segments) coincide with landscape objects (patches)." Unfortunately, the same Size-Constrained Region Merging (SCRM) algorithm proposed by Castilla *et al*. makes no exception to their criticism since its "correspondence between radiometric similarity and semantic similarity is not straightforward" (Castilla et al., 2008).

To summarize, according to Castilla *et al*. the conceptual framework of OBIA requires generation of symbolic image segments as output. This is the same claim made by cognitive psychology (see this text above) (Corcoran & Winstanley, 2007; Corcoran et al., 2010; Vecera & Farah, 1997). This also agrees with Marr's statement: "vision goes symbolic immediately, right at the level of zero-crossing (primal sketch)... without loss of information" (Marr, 1982)

(p. 343), refer to Part I Section 2.3. As a consequence, if this conjecture holds, then existing commercial image segmentation algorithms, whose claim is to be at the basis of the GEOBIA success (Definiens Imaging GmbH, 2004; Esch et al., 2008), are actually in contrast with the true conceptual framework of GEOBIA, which requires detection of semantic image segments (e.g., landscape objects or patches).

Unfortunately, in spite of the aforementioned contributions found in existing literature, most members of the CV and RS communities, including Diamant (Diamant, 2005; Diamant, 2008; Diamant, 2010a; Diamant, 2010b) (refer to Part I Section 2.5), appear to ignore the inherently ill-posed (subjective) nature of the image segmentation (region extraction/ contour detection) problem. As a consequence, literally dozens of "novel" segmentation (region extraction/contour detection) algorithms are published each year (Zamperoni, 1996). For example, due to the availability of a commercial GEOBIA software developed by a German company (Definiens Imaging GmbH, 2004; Esch et al., 2008), OBIA approaches are currently considered the state-of-the-art in both scientific and commercial RS image mapping applications (Castilla et al**.**, 2008; Hay & Castilla, 2006).

In commercial GEOBIA systems, to reduce the number of empirical segmentation parameters (Esch et al., 2008), a multi-scale (hierarchical) iterative segmentation first stage is employed (Definiens Imaging GmbH, 2004). As output, a hierarchical segmentation algorithm generates multi-scale segmentation solutions in the hope that the target image will appear correctly segmented at some scale. However, quantitative multi-scale assessment of segmentation quality indices requires ground truth data at each scale which are impossible or impractical to obtain in RS common practice (Corcoran & Winstanley, 2007). Therefore, the "best" segmentation map must be selected by the user on an *a posteriori* basis from the available set of multi-scale segmentation solutions according to heuristic, subjective and/or qualitative criteria analogous to those employed in the selection of prior segmentation parameters. In practice, exploitation of a hierarchical segmentation algorithm does not make a driven-without-knowledge segmentation first stage easier to use. In addition, hierarchical segmentation algorithms are computationally intensive and require large memory occupation.

The conclusion is that, to date, in spite of its commercial success, GEOBIA remains affected by a lack of general methodological consensus and research (Hay & Castilla, 2006). Scientific disagreement on the conceptual framework of GEOBIA finds its origin in the well-known information gap existing between physical information (sensations) and semantic information (percepts) (Matsuyama & Shang-Shouq Hwang, 1990) (see Part I Section 2.2.2 and Part I Section 2.3). Since GEOBIA appears unable to generate semantic image segments (e.g., landscape objects) in the pre-attentive vision phase, it appears unsuitable for filling the information gap between raster sub-symbolic imagery and vector symbolic geospatial information (typically dealt with by geographic information systems, GIS).

2.4.2 Labeled data learning for classification and function approximation

Labeled (supervised) data learning approaches deal with either classification or function approximation (regression) problems whose output variables are discrete semantic and continuous non-semantic respectively, see Fig. 1 (Alpaydin, 2010; Bishop, 1995; Cherkassky & Mulier, 2006; Mather, 1994; Mitchell, 1997).

In classification problems where the available training data set is assumed to be fully reliable (which may not always be the case (Bruzzone & Persello, 2009)), the goal of a classifier capable of learning from labeled data is to achieve a perfect fit of the training data set (to reduce to zero the training error) and, at the same time, make good semantic predictions for new (previously unobserved) inputs (to reduce to zero the testing error). An adaptive classifier can be trained in various ways, namely, on-line (sequential learning (Bishop, 1995), stochastic learning (Cherkassky & Mulier, 2006), when a large or infinite input data sequence is available and/or real-time adaptation is required), batch (it requires the storage of a complete and finite training data set (Bishop, 1995)) and semi-batch (Wilson & Martinez, 2000). In addition, there are many statistical classifiers. The most widely used statistical classifiers are the plug-in parametric maximum likelihood (ML) classifier, the non-parametric Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) networks, kernel methods (also called memory-based, which require the storage of a complete data set (Mitchell, 1997)) such as the SVM and the k-nearest neighbor (K-NN) algorithm, the naive Bayes classifier, adaptive (statistical) decision-trees such as the Classification And Regression Tree (CART), adaptive rule-based systems, mixture of experts (Jordan & Jacobs, 1994), etc. (Alpaydin, 2010; Bishop, 1995; Cherkassky & Mulier, 2006; Duda et al., 2001; Mitchell, 1997).

Classifier performance depends greatly on the characteristics of the labeled data set to be classified (Baraldi et al., 2006b). In other words, there is no single classifier that works best on all given problems; this is also referred to as the "no free lunch" theorem. In practical contexts, classification model selection, i.e., determining a suitable classifier for a given problem, is still more an art than a science.

In reinforcement learning the agent is rewarded for good responses and punished for bad ones. These can be analyzed in terms of decision theory, using concepts such as utility (Cherkassky & Mulier, 2006).

Function regression (curve fitting) takes a finite set of numerical continuous input-output pair samples and attempts to discover an unknown continuous (smooth) deterministic function which, together with added Gaussian noise, would generate those target outputs from the inputs (Bishop, 1995). The goal of function approximation is not to learn an exact representation (interpolation) of the training data, but rather to build a statistical model of the physical process that generates the training labeled data. This statistical model ought to be capable of the best trade-off between: (a) achieving a good fit of the training data (to keep low the bias term of a sum-of-squares error function) and (b) obtaining a reasonably smooth function that is not over-fitted to the training data (to keep the variance term of a sum-of-squares error function low). This is important if the self-organizing (adaptive) function approximation system is to exhibit good generalization, i.e., to make good numerical predictions for new (previously unobserved) inputs (Bishop, 1995).

To summarize, to properly deal with discrete semantic or continuous non-semantic output values, labeled (supervised) data learning systems feature different functional hypotheses and properties. For example:

- they adopt different cost functions, namely, the cross-entropy error function for adaptive classifiers versus the sum-of-squares error for function approximation approaches (Bishop, 1995) (p. 230).

- When the training labeled data set is assumed to be fully reliable the goal of adaptive classifiers is to reduce to zero both training and testing errors (e.g., if the training error is equal to zero then a classifier is called consistent (Baraldi & Alpaydin, 2002b; Mitchell, 1997)). Vice versa, reducing to zero the bias term in function regression is not recommended because it would imply over-fitting to the training data assumed to be inherently affected by Gaussian noise (which is not the case for exact interpolators) (Bishop, 1995).

2.5 Diamant's image segmentation and contour detection algorithms as proofs of his concepts

As proofs of his concepts (see Part I Section 2.2.3) Diamant presents an image segmentation algorithm and a contour detection algorithm which are summarized below.

2.5.1 Multi-scale image segmentation algorithm

In (Diamant, 2005), a multi-scale image segmentation algorithm is presented and applied to a toy problem, namely, a panchromatic (one-band) image of 640 × 480 pixels in size. The proposed segmentation algorithm is as follows.

1. Low-pass (smoothing) dyadic (sub-sampling by a factor of 2) image decomposition (down-scaling). Image decomposition levels are identified with integer numbers l = 0,..., L, L+1, where level 0 identifies the input image at full spatial resolution. Value L > 0 is set to 4, thus the maximum down-scale level is L+1 = 5. A simple dyadic multi-scale panchromatic (one-band) image decomposition and averaging operator is applied as follows.

$$g^{l+1}(x,y) = [g^{l}(2x,2y) + g^{l}(2x + 1,2y) + g^{l}(2x + 1,2y + 1) + g^{l}(2x,2y + 1)]/4, \quad l= 0, ..., L > 0, \tag{1-1}$$

   where $g^{l+1}(x,y)$ is the gray-level value of a (down-scaled parent) pixel at the (x,y) coordinate position in a higher (l+1)-level image while $g^{l}(2x,2y)$ and its three nearest neighbors listed in Eq. (1-1) are the corresponding (up-scaled children) pixels within an image array at the lower level l.

2. Single-scale image segmentation algorithm run at the top (coarsest) (L+1)-level of the decomposition pyramid. Diamant claims that since the image size at the top level of the pyramid is significantly reduced and a severe data averaging is attained, any well-known segmentation methodology would suffice. Diamant's proprietary segmentation technique firstly outlines image boundaries (contours) (see Part I Section 2.4.1.2). Secondly, contiguous pixels of "similar" appearance (based on an unknown similarity measure and decision rule) within non-closed contours are aggregated in spatially connected segments (this is apparently a region growing from non-closed contours approach, e.g., refer to (Baraldi & Parmiggiani, 1995)). Thirdly, the segment-based mean intensity image, called characteristic intensity, is computed (this is a piecewise constant image approximation of the input image generated by replacing every pixel with the mean value of the segment where that pixel is located).

3. (Coarse-to-fine spatial resolution) mean image and segmentation map up-scaling. At each level l = L + 1, ..., 1, with step -1, the mean image and the segmentation map are expanded to the size of the image at the nearest lower level (l-1) (at finer spatial resolution). The expansion rule is simple and the same for both up-scaling operations: the value of each parent pixel at level l is assigned to its four children at level (l-1). Diamant claims that since image regions feature a low inter-segment intensity variability, the majority of newly assigned pixels are determined in a sufficiently correct manner. Only pixels lying on object boundaries or seeds of newly emerging objects can significantly deviate from their up-scaled assigned value. Taking the corresponding l-level of the down-scaled image as a reference, these pixels can easily (!?) be detected and subjected to a refinement cycle. Here they are allowed to adjust themselves to the ''proper'' nearest neighbors, which certainly belong to one of the previously labeled regions or to the newly emerging ones. Unlike the lossless image decomposition/reconstruction procedure provided by Burt and Adelson's Gaussian/Laplacian pyramid (Burt & Adelson, 1983), in the Diamant case the exact reconstruction of an image is not required. In Diamant's opinion "only (?!) in special cases - medical, scientific, military, fine-art, and a couple (!?) of other applications - the reconstruction fidelity of the original image can be critically important" (Diamant, 2005), which is to say it is critical in all quantitative rather than qualitative CV applications! For example, RS image understanding applications require small, but genuine image details, say, roads, to be well preserved, which is tantamount to saying that RS image applications are among the "couple (!?) of other applications" where high fidelity in multi-scale encoding (decomposition)/decoding (reconstruction) is required.

A critical analysis of the Diamant image segmentation algorithm can be found in Part II Section 3.1.

2.5.2 Single-scale image contour detection algorithm

In (Diamant, 2005) Diamant presents a single-scale image contour detection algorithm and applies it to a toy problem, namely, a panchromatic image 256 × 256 pixels in size. This contour detector provides a measure of local information, $I_{loc}(x,y)$, as a product of two terms.

$$I_{loc}(x,y) = I_{int}(x,y) \times I_{top}(x,y) \tag{1-2}$$

where (x,y) are the central pixel coordinates in a (2-D) image array, factor $I_{int}(x,y)$ is the intensity change component and factor $I_{top}(x,y)$ is considered a measure of topological confidence (uncertainty). In Eq. (1-2) term $I_{int}(x,y)$ is estimated as follows.

$$I_{int}(x,y) = \frac{1}{8}\sum_{n=1}^{8}\left|g_c(x,y) - g_n(x,y)\right| \geq 0. \tag{1-3}$$

Thus, in Eq. (1-2) the first term $I_{int}(x,y)$ is estimated as the mean absolute difference between the central pixel gray value, $g_c(x,y)$, and the gray levels of its 8-adjacency neighbors, $g_n(x, y)$, n = 1, ..., 8.

In Eq. (1-2) the second term $I_{top}(x,y)$ is computed in two steps. Firstly, an expression for a pixel's interrelationship with its surrounding is defined as follows.

$$\text{status}(x,y) = 8g_c(x, y) - \sum_{n=1}^{8} g_n(x, y). \quad (1\text{-}4)$$

It is worthy of note that status(x, y) is equivalent to a contrast value computed by an isotropic *mexican-hat* operator centered on pixel (x, y). The shortest status(x, y) description (encoding) would be in a binary form, for example, 0 if status is negative, and 1 otherwise. Status(x, y) is evaluated for every pixel (x, y) in an image and mapped into a binary status map of the same size as the input image. Secondly, the spatial (topological) interactions of a pixel with its 8-adjacency neighbors can be estimated using the binary status map:

$$I_{top}(x,y) = p\,(1 - p) = (m/8)\,[(8 - m)/8],\ m \in \{0, 8\}, \quad (1\text{-}5)$$

where p is the probability that the central pixel and its surrounding ones share the same status, such that $m \in \{0, 8\}$ is the number of 8-adjacency pixels that share the same status with the central pixel in the 2-D array position (x, y). Any $I_{top}(x, y)$ value is computed for every pixel (x, y) and saved in a special image of the size of the input image.

Diamant considers peaks (local extrema) in $I_{loc}(x,y)$ = Eq. (1-2) = $I_{int}(x,y) \times I_{top}(x,y)$ = Eq. (1-3) × Eq. (1-5) as signs of a visible edge present at a given location. However, establishing a proper threshold for local extrema has always been a hard and sophisticated matter. To overcome this difficulty, Diamant proposes to gather a cumulative histogram of I_{loc} values. At first, a number of equal intervals (bins) is selected and a histogram (first-order statistic) of the I_{loc} image is constructed in sequence for every histogram bin as follows: if the pixel-based I_{loc} value is greater than or equal to the bin's lower bound, then this bin counter is increased by one. As a result, the first bin represents the cardinality of all I_{loc} values > 0. It is now explicitly visible what part of the whole ''image information content'' is carried out by I_{loc} values equal to or greater than a particular bin lower bound. This can be used as a (subjective!) threshold for appropriate image point assignment (marking). In such a way, a set of different information content-related thresholds can be established, which can address diversified task-related requirements. For example, the most prominent image points are marked in dark gray, carrying more than 50% of the whole information content. Less important image parts can be marked in half-gray, carrying between 50 and 70% of information content, and the lowest importance image parts are marked in light gray, carrying 70 to 85% residuals of the information content. The proposed image point marking technique can be effectively used to create more enhanced low-level information content descriptors. For example, based on the status image generated from Eq. (1-4), an edge-localization image can be displayed where dark-gray is assigned to the lower intensity sides of the edges and light-gray to the higher intensity edge sides (Diamant, 2005).

A critical analysis of the Diamant image contour detection algorithm can be found in Part II Section 3.3.

2.6 Four levels of understanding of an RS-IUS

It is important to remember that there are four levels of analysis (understanding) of any information processing device, including RS-IUSs. They are listed below (Baraldi et al., 2010b; Baraldi, 2011a; Marr, 1982).

i. Computational theory (system architecture). According to Marr, the linchpin of success in attempting to solve the CV problem is that of addressing the computational theory rather than algorithms or implementations (Marr, 1982). In other words, if the vision device architecture is inadequate, even sophisticated algorithms can produce low-quality outputs. On the contrary, improvement in the vision system architecture might achieve twice the benefit with half the effort (which is an adaptation of the original words by Wang (Fangju Wang, 1990)). For example, a two-stage stratified hierarchical hybrid RS-IUS architecture (see Part II Fig. 3) has been proposed in recent literature (Baraldi et al., 2006a; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi et al., 2010c; Baraldi, 2011a; Baraldi, 2011b), as an alternative to the current state-of-the-art two-stage GEOBIA architecture, hereafter referred to as two-stage segment-based hybrid RS-IUS architecture (see Part II Fig. 2).
ii. Knowledge/information representation. According to Wang, "if knowledge representation is poor, even sophisticated algorithms can produce inferior outputs. On the contrary, improvement in representation might achieve twice the benefit with half the effort" (Fangju Wang, 1990). For example, in (Baraldi et al., 2010c; Baraldi, 2011b) a crisp-to-fuzzy SIAM™ transition has been accomplished to model class mixtures.
iii. Algorithm design. This level deals with the design of the algorithm selected to fill each of the data processing modules comprised in the system architecture (refer to point (i) above). According to (Page-Jones, 1988), structured system design is "everything but code".
iv. Implementation. This level deals with the source code generation for every algorithm designed at point (iii) above.

2.7 Quality Assurance Framework for EO (QA4EO)

Delivered by the Working Group on Calibration and Validation (WGCV) of the Committee of Earth Observations (CEOS), the space arm of the Group on Earth Observations (GEO) (GEO, 2005; GEO, 2008b), the QA4EO guidelines (GEO/CEOSS, 2008) consider mandatory the following actions: (i) calibration and validation (Cal/Val) activities from sensor build to end-of-life and (ii) every sensor-derived data product must be provided with metrological/ statistically-based quality indicators (QIs) featuring a degree of uncertainty in measurement. Unfortunately, in RS common practice, these international guidelines are often ignored by scientists, practitioners and whole institutions (Baraldi, 2009).

2.7.1 Calibration and validation (Cal/Val) activities from sensor build to end-of-life

QA4EO considers mandatory an appropriate coordinated program of Cal/Val activities throughout all stages of a spaceborne mission, from sensor build to end-of-life (GEO/CEOSS, 2008). This ensures the harmonization and interoperability of multi-source observational data and derived products required by international programs such as the on-going GEOSS and GMES projects (GEO, 2008b; GEO, 2005) (refer to Part I Section 1).

In spite of the QA4EO recommendations and although it is regarded as common knowledge in the RS community, *radiometric calibration*, i.e., the transformation of dimensionless digital numbers (DNs) into a physical unit of measure related to a community-agreed radiometric scale, is often neglected in literature and surprisingly ignored by scientists, practitioners and

institutions involved with RS common practice including large-scale spaceborne image mosaicking and mapping (Baraldi et al., 2006a; Baraldi, 2009; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi, 2011a).

A relevant extension of the QA4EO recommendation for radiometric calibration of multi-source EO data is the following.

"Radiometric calibration not only ensures the harmonisation and interoperability of multi-source observational data according to the QA4EO guidelines, but is a necessary, although insufficient, condition for automating the quantitative analysis of EO data" (Baraldi et al., 2006a; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi, 2011a) in RS data understanding problems other than toy problems at small data scale and coarse semantic granularity. By definition, a data processing system is *automatic* when it requires no user-defined parameter to run, therefore its user-friendliness cannot be surpassed (refer to Part I Section 2.8).

This necessary condition for automatic EO data understanding agrees with common sense, summarized by the expression: "garbage in means garbage out". In the terminology of MAL and CV, the radiometric calibration constraint augments the degree of prior knowledge of a RS-IUS required to complement the intrinsic insufficiency (ill-posedness) of (2-D) image features, i.e., radiometric calibration makes the inherently ill-posed CV problem better posed (Baraldi et al., 2010a; Baraldi, 2011a; Matsuyama & Shang-Shouq Hwang, 1990).

To summarize, in disagreement with the QA4EO guidelines, most existing scientific and commercial RS-IUSs, such as those listed in Table 1, do not require RS images to be radiometrically calibrated and validated. As a consequence, according to the aforementioned necessary condition for automating the quantitative analysis of EO data, these RS-IUSs are semi-automatic and/or site-specific (since one scene may represent, say, apples, while any other scene, even if contiguous or overlapping, may represent, say, oranges), refer to Table 1. Secondly, Table 1 shows that unlike SIAM™, the ERDAS Atmospheric Correction for satellite imagery (ATCOR3) (Richter, 2006) requires as input an MS image radiometrically calibrated into surface reflectance values exclusively. This implies that the ERDAS ATCOR3 software considers mandatory the inherently ill-posed and difficult-to-solve MS image atmospheric correction pre-processing stage which requires user intervention to make it better posed (Baraldi, 2011a). Thus, unlike SIAM™, the ERDAS ATCOR3 satisfies the necessary condition for automating the quantitative analysis of EO data, but is insufficient to provide a RS image classification problem with an automatic workflow requiring no user-defined empirical parameter to be based on heuristic criteria.

2.7.2 Quality Indicators (QIs) with a degree of uncertainty

In addition to considering mandatory an appropriate coordinated program of Cal/Val activities throughout all stages of a spaceborne mission, from sensor build to end-of-life (see Section 2.7.1), the QA4EO guidelines require that every sensor-derived data product generated across a satellite-based measurement system's processing chain be provided with metrological/ statistically-based QIs featuring a degree of uncertainty in measurement (GEO/CEOSS, 2008). Unfortunately, in RS common practice, as well as in existing literature,

Commercial RS-IUSs	Sub-symbolic (asemantic) versus symbolic (semantic) information primitives, namely, pixels / (2-D) objects (regions, segments) / strata	Radiometric calibration (RAD. CAL.) requirement according to the international QA4EO guidelines
PCI Geomatics GeomaticaX	Sub-symbolic pixels	NO RAD. CAL. Þ semi-automatic and site-specific
eCognition Server by Definiens AG	Unsupervised data learning sub-symbolic objects	NO RAD. CAL. Þ semi-automatic and site-specific
Pixel- and Segment-based versions of the Environment for Visualizing Images (ENVI) by ITT VIS	Either sub-symbolic pixels or unsupervised data learning sub-symbolic objects	NO RAD. CAL. Þ semi-automatic and site-specific
ERDAS IMAGING Objective	Supervised data learning symbolic objects	NO RAD. CAL. Þ semi-automatic and site-specific
ERDAS Atmospheric Correction-3 (ATCOR3) (Richter, 2006)	Sub-symbolic pixels	Consistent with the QA4EO recommendations: surface reflectance, SURF Þ inherently ill-posed atmospheric correction first stage Þ semi-automatic and site-specific.
Novel two-stage stratified hierarchical RS-IUS employing SIAM™ as its preliminary classification first stage	Prior knowledge-based symbolic pixels $\in$ symbolic objects $\in$ symbolic strata	Consistent with the QA4EO recommendations: top-of-atmosphere (TOA) reflectance (TOARF) or surface reflectance (SURF) values, with TOARF $\supseteq$ SURF $\Rightarrow$ atmospheric correction is optional. Automatic and robust to changes in RS optical imagery acquired across time, space and sensors.

Table 1. Existing commercial RS-IUSs and their degree of match with the international QA4EO quidelines.

these international guidelines are often ignored by scientists, practitioners and whole institutions (Baraldi, 2009). For example, most works published in RS literature assess and compare spaceborne image classification algorithms in terms of mapping accuracy exclusively, which corresponds to only one of several operational QIs of a RS-IUS (refer to Part I Section 2.8). Moreover, these classification accuracy estimates are rarely provided with a degree of uncertainty in measurement. This violates well-known laws of sample statistics (Congalton & Green, 1999; Foody, 2002; Jain et al., 2000), together with common sense envisaged under the international guidelines of the QA4EO (GEO/CEOSS, 2008).

It is well known, but often forgotten in common practice that any evaluation measure is inherently non-injective (Baraldi, 2011a). For example, in classification map accuracy assessment and comparison, different classification maps may produce the same confusion matrix while different confusion matrices may generate the same confusion matrix accuracy measure, such as overall accuracy. These observations suggest that *no single universally acceptable measure of quality, but instead a variety of quality indices, should be employed in practice* (Congalton & Green, 1999; Foody, 2002). To date, this general conclusion is neither obvious nor community-agreed. For example, this conclusion implies that when a test image and a reference (original) image pair is given, common attempts to identify a unique (universal) reliable image quality index, such as the relative dimensionless global error ERGAS proposed in (Wald et al., 1997), the universal image quality index Q (Wang & Bovik, 2002), the global image quality measure Q4 (Alparone et al., 2004), and the quality index with no reference QNR (Alparone et al., 2006), are inherently undermined as contradictions in terms.

In recent years the issue of uncertainty in spatial data has become increasingly recognized by the RS and geographic information systems (GIS) communities (Friedl et al., 2001). Spatial uncertainty analysis investigates sources of inaccuracies in geospatial data acquisition and understanding and investigates error propagation through a RS (2-D) image processing chain. For example, post-classification change detection between two classification maps of overall accuracy $OA_1 \in [0, 1]$ and $OA_2 \in [0, 1]$, respectively, features a change detection OA (COA) such that $COA \leq (OA_1 \times OA_2)$ (Lunetta & Elvidge, 1999). For example, Friedl *et al*. identify three primary sources of errors in spatial information generated from RS imagery (Friedl et al., 2001).

1. Errors introduced through the image acquisition process (e.g., spectral and spatial image distorsion).
2. Errors produced by the application of image processing techniques, namely, (a) image pre-processing algorithms (e.g., atmospheric correction, geometric correction, radiometric calibration) and (b) image understanding techniques (e.g., spatial and semantic accuracies in classification mapping).
3. Errors associated with interactions between the instrument time, spatial and spectral resolution and the physical nature and scale of an ecological process on the ground (e.g., pixels affected by class mixture).

2.8 Operational Quality Indicators (QIs) of an RS-IUS

In operational contexts a RS-IUS is defined as a low performer if at least one among several operational QIs scores low. Typical operational qualities of a RS-IUS encompass the following (Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi, 2011a).

i. Degree of automation. For example, a data processing system is *automatic* when it requires no user-defined parameter to run, therefore its user-friendliness cannot be surpassed. When a data processing system requires neither user-defined parameters nor reference data samples to run, it is termed "*fully automatic*" (Qiyao Yu & Clausi, 2007).
ii. Effectiveness, e.g., classification accuracy and spatial accuracy (Baraldi et al., 2005; Persello & Bruzzone, 2010).

iii. Efficiency, e.g., computation time, memory occupation.
iv. Economy (costs). Related to manpower and computing power. For example, open source solutions are welcome to reduce costs of software licenses. Supervised data learning approaches (e.g., SVMs, OBIA systems, etc.) require reference training samples which are typically scene-specific, expensive, tedious, difficult or impossible to collect.
v. Robustness to changes in the input data set, e.g., changes due to noise in the data.
vi. Robustness to changes in input parameters, if any exist.
vii. Maintainability / scalability / re-usability to keep up with changes in users' needs and sensor properties.
viii. Timeliness, defined as the time span between data acquisition and product delivery to the end user. It increases monotonically with manpower, e.g., the manpower required to collect site-specific training samples.

The aforementioned list of operational QIs is neither irrelevant nor obvious. For example, a low score in operational QIs may explain why the literally hundreds of so-called novel low-level (sub-symbolic) and high-level (symbolic) image processing algorithms presented each year in scientific literature typically have a negligible impact on commercial RS image processing software (Zamperoni, 1996). This conjecture is consistent with the fact that most works published in RS literature assess and compare spaceborne image classification algorithms in terms of mapping accuracy exclusively, which corresponds to the sole operational performance indicator (ii) listed above. Moreover, these classification accuracy estimates are rarely provided with a degree of uncertainty in measurement. This violates well-known laws of sample statistics (Congalton & Green, 1999; Foody, 2002; Jain et al., 2000), together with common sense envisaged under the international guidelines of the QA4EO (see Part I Section 2.7.2) (GEO/CEOSS, 2008).

3. Conclusions

The goal of this work is to revise, integrate and enrich previous analyses found in related papers about recent developments in the design and implementation of an operational automatic multi-sensor multi-resolution near real-time two-stage hybrid stratified hierarchical RS-IUS (Baraldi et al., 2006a; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi, 2011a).

For publication reasons this work is split into Part I and Part II. In Part I Section 2, related works, concepts and definitions are revised to provide this paper with a significant survey value and make it self-contained. In Part II Section 2, the survey of past works is completed. The original contribution of this work can be found in Part II Section 3 to Part II Section 7.

4. Acknowledgments

This material is partly based upon work supported by the National Aeronautics and Space Administration under Grant/Contract/Agreement No. NNX07AV19G issued through the Earth Science Division of the Science Mission Directorate. The research leading to these results has also received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement n° 263435. This author wishes to thank the Editorial Board of InTech for its competence and willingness to help.

5. References

Acton, S.T. & Landis, J. (1997). Multispectral anisotropic diffusion. *Int. J. Remote Sensing*, Vol. 18, pp. 2877-2886.

Alparone, L.; Baronti, S.; Garzelli, A. & Nencini, F. (2004). A global quality measurement of pan-sharpened multispectral imagery. *IEEE Geosci. Remote Sensing Letters*, Vol. 1, No. 4, pp. 313-317.

Alparone, L.; Aiazzi, B.; Baronti, S.; Garzelli, A. & Nencini, F. (2006). A new method for MS+Pan image fusion assessment without reference, *IEEE International Geoscience and Remote Sensing Symposium* (IEEE IGARSS 2006), Denver, Colorado, Jul. 31–Aug. 4. 2006.

Alpaydin, E. (2010). *Introduction to Machine Learning (Adaptive Computation and Machine Learning*. The MIT Press, Cambridge, MT.

Baatz, M.; Hoffmann, C.; Willhauck, G. Progressing from object-based to object-oriented image analysis. In Object-Based Image Analysis–Spatial Concepts for Knowledge-driven Remote Sensing Applications; Blaschke, T., Lang, S., Hay, G.J., Eds.; Springer-Verlag: New York, NY, 2008, Chapter 1.4, pp. 29- 42.

Backer, E. & Jain, A. K. (1981). A clustering performance measure based on fuzzy set decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. PAMI-3, No. 1, pp. 66–75.

Baraldi, A. & Parmiggiani, F. (1995). A refined Gamma MAP SAR speckle filter with improved geometrical adaptivity. *IEEE Trans. Geosci. Remote Sensing*, Vol. 33, No. 5, pp. 1245-1257.

Baraldi, A. & Parmiggiani, F. (1996a). Combined detection of intensity and chromatic contours in color images. *Optical Engineering*, Vol. 35, No. 5, pp. 1413-1439.

Baraldi, A. & Parmiggiani, F. (1996b). Single linkage region growing algorithms based on the vector degree of match. *IEEE Trans. Geosci. Remote Sensing*, Vol. 34, No. 1, pp. 137-148.

Baraldi, A. & Blonda, P. (1999a). A survey of fuzzy clustering algorithms for pattern recognition: Part I. *IEEE Trans. Systems, Man and Cybernetics - Part B: Cybernetics*, Vol. 29, No. 6, pp. 778-785.

Baraldi, A. & Blonda, P. (1999b). A survey of fuzzy clustering algorithms for pattern recognition: Part II. *IEEE Trans. Systems, Man and Cybernetics - Part B: Cybernetics*, Vol. 29, No. 6, pp. 786-801.

Baraldi, A. & Alpaydin, E. (2002a). Constructive feedforward ART clustering networks—Part I. *IEEE Trans. Neural Netw.*, Vol. 13, No. 3, pp. 645–661.

Baraldi, A. & Alpaydin, E. (2002b). Constructive feedforward ART clustering networks—Part II. *IEEE Trans. Neural Netw.*, Vol. 13, No. 3, pp. 662–677.

Baraldi, A.; Bruzzone, L. & Blonda, P. (2005). Quality assessment of classification and cluster maps without ground truth knowledge. *IEEE Trans. Geosci. Remote Sensing*, Vol. 43, No. 4, pp. 857-873.

Baraldi, A.; Puzzolo, V.; Blonda, P.; Bruzzone, L. & Tarantino, C. (2006a). Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images, *IEEE Trans. Geosci. Remote Sensing.* Vol. 44, No. 9, pp. 2563-2586.

Baraldi, A.; Bruzzone, L.; Blonda, P. & Carlin, L. (2006b). Badly-posed classification of remotely sensed images - An experimental comparison of existing data labeling systems, *IEEE Trans. Geosci. Remote Sensing*, Vol. 44, No. 1, pp. 214-235.

Baraldi, A. (2009). Impact of radiometric calibration and specifications of spaceborne optical imaging sensors on the development of operational automatic remote sensing image understanding systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 2, No. 2, pp. 104-134.

Baraldi, A.; Durieux, L.; Simonetti, D.; Conchedda, G.; Holecz, F. & Blonda, P. (2010a). Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/ OrbView/GeoEye and DMC/SPOT-1/-2 imagery – Part I: System design and implementation. *IEEE Trans. Geosci. Remote Sensing*, Vol. 48, No. 3, pp. 1299 - 1325.

Baraldi, A.; Durieux, L.; Simonetti, D.; Conchedda, G.; Holecz, F. & Blonda, P. (2010b). Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/ OrbView/GeoEye and DMC/SPOT-1/-2 imagery – Part II: Classification accuracy assessment. *IEEE Trans. Geosci. Remote Sensing*, Vol. 48, No. 3, pp. 1326 - 1354.

Baraldi, A.; Wassenaar, T. & Kay, S. (2010c). Operational performance of an automatic preliminary spectral rule-based decision-tree classifier of spaceborne very high resolution optical images. *IEEE Trans. Geosci. Remote Sensing*, Vol. 48, No. 9, pp. pp. 3482 - 3502.

Baraldi, A. Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA): Levels of understanding and degrees of novelty of an operational automatic two-stage stratified hierarchical hybrid remote sensing image understanding system, Remote Sens., submitted for consideration for publication, rs-10905, 2011.

Baraldi, A. (2011b). Fuzzification of a crisp near real-time operational automatic spectral rule-based decision-tree preliminary classifier of multi-source multi-spectral remotely-sensed images, *IEEE Trans. Geosci. Remote Sensing*, accepted for publication, July 2011.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, United Kingdom.

Bruzzone, L. & Carlin, L. (2006). A multilevel context-based system for classification of very high spatial resolution images. IEEE *Trans. Geosci. Remote Sens.*, Vol. 44, No. 9, pp. 2587–2600.

Bruzzone, L. & Persello, C. (2009). A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples. *IEEE Trans. Geosci. Remote Sensing*, Vol. 47, No. 7, pp. 2142-2154.

Burr, D. C. & Morrone, M. C. (1992). A nonlinear model of feature detection, In: *Nonlinear Vision: Determination of Neural Receptive Fields, Functions, and Networks*, R. B. Pinter & N. Bahram, (Eds.), pp. 309-327, CRC Press, Boca Raton, Florida.

Burt, P. & Adelson, E. (1983). The Laplacian pyramid as a compact image code. *IEEE Trans. Communications*, Vol. COM-31, No. 4, pp. 532-540.

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 8, pp. 679-714.

Carson, C.; Belongie, S.; Greenspan, H. & Malik, J. (1997). Region-Based Image Querying, *Proc. Int'l Workshop Content-Based Access of Image and Video libraries*, San Juan , Puerto Rico, June 20, 1997.

Castilla, G., Hay, G. J. & Ruiz-Gallardo, J. R. (2008). Size-constrained Region Merging (SCRM): An automated delineation tool for assisted photointerpretation. *Photogramm. Eng. Remote Sens*., Vol. 74, No. 4 (Apr. 2008), pp. 409-429.

Chengquan Huang; Kuan Song; Sunghee Kim; Townshend, J.; Davis, P.; Masek, J. G. & Goward, S. N. (2008). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, Vol. 112, pp. 970–985.

Cherkassky , V. & Mulier, F. (2006). *Learning From Data: Concepts, Theory, and Methods.* Wiley, New York.D'Elia, S. (2009). European Space Agency (ESA), personal communication.

Congalton, R. G. & Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data*. Lewis Publishers, Boca Raton, Florida.

Corcoran, P. & Winstanley, A. (2007). Using texture to tackle the problem of scale in landcover classification, In: *Object-Based Image Analysis - Spatial concepts for knowledge-driven remote sensing applications*, T. Blaschke, S. Lang & G. Hay, (Eds.), pp. 113-132, Springer Lecture Notes in Geoinformation and Cartography.

Corcoran, P.; Winstanley, A. & Mooney, P. (2010). Segmentation performance evaluation for object-based remotely sensed image analysis. *Int. J. Remote Sensing*, Vol. 31, No. 3, pp. 617-645.

Definiens Imaging GmbH. (2004). *eCognition* User Guide 4.

Delves, L.; Wilkinson, R.; Oliver, C. & White, R. (1992). `Comparing the performance of SAR image segmentation algorithms,. *Int. J. Remote Sensing*, Vol. 13, No. 11, pp. 2121-2149.

Diamant, E. (2005). Searching for image information content, its discovery, extraction, and representation. *Journal of Electronic Imaging*, Vol. 14, No. 1, pp. 1-11.

Diamant, E. (2008). I'm Sorry to Say, But Your Understanding of Image Processing Fundamentals Is Absolutely Wrong, In: *Brain, Vision and AI*, C. Rossi, (Ed.), Chapter 5, pp. 95-110, In-Tech Publishing.

Diamant, E. (2010a). Not only a lack of a right definition: Arguments for a shift in information-processing paradigm. 17.04.2011, Available from: http://arxiv.org/ftp/arxiv/papers/1009/1009.0077.pdf

Diamant, E. (2010b). Machine Learning: When and Where the Horses Went Astray?, In: *Machine Learning*, Yagang Zhang, (Ed.), pp. 1-18, In-Tech Publishing. 17.04.2011, Available from: http://arxiv.org/abs/0911.1386

Duda, R.; Hart, P. & Stork, D. (2001). *Pattern Classification*. Wiley, New York.

Esch, T.; Thiel, M.; Bock, M.; Roth, A. & Dech, S. (2008). Improvement of image segmentation accuracy based on multiscale optimization procedure. *IEEE Geosci. Remote Sensing Letters*, Vol. 5, No. 3 (Jul. 2008), pp. 463-467.

European Space Agency (ESA). (2008). GMES Observing the Earth, 17.04.2011, Available from: http://www.esa.int/esaLP/SEMOBSO4KKF_LPgmes_0.html

Fangju Wang. (1990). Fuzzy supervised classification of remote sensing images. *IEEE Trans. Geosci. Remote Sensing*, Vol. 28, No. 2, pp. 194-201.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, Vol. 80, pp. 185-201.

Friedl, M.A.; McGwire, K.C. & McIver, D. K. (2001). An overview of uncertainty in optical remotely sensed data for ecological applications, In: *Spatial Uncertainty in Ecology:*

Implications for Remote Sensing and GIS Applications, C.T. Hunsaker, M.F. Goodchild, M.A. Friedl & T.J. Case, (Eds), pp. 258–283, Springer, New York.

Fritzke, B. (1997). *Some competitive learning methods* (Draft document), 17.04.2011, Available from: http://www.neuroinformatik.ruhr-unibochum.de/ini/VDM/research/gsn/DemoGNG.

GEO/CEOSS. (2008). A Quality Assurance Framework for Earth Observation, Version 2.0, 17.04.2011, Available from: http://calvalportal.ceos.org/CalValPortal/showQA4EO.do?section=qa4eoIntro

Global Monitoring for Environment and Security (GMES) (2011). 17.04.2011, Available from: http://www.gmes.info

Gouras, P. (1991). Color vision, In: *Principles of Neural Science*, E. Kandel & J. Schwartz, (Eds.), pp. 467-479, Appleton and Lange, Norwalk, Connecticut.

Group on Earth Observations (GEO). (2005). The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan, 17.04.2011, Available from: http://www.earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf

Group on Earth Observations (GEO). (2008a). GEO announces free and unrestricted access to full Landsat archive, 17.04.2011, Available from: www.fabricadebani.ro/userfiles/GEO_press_release.doc

Group on Earth Observations (GEO). (2008b). GEO 2007-2009 Work Plan: Toward Convergence, 17.04.2011, Available from: http://earthobservations.org

Gutman, G. *et al*., (Eds.). (2004). *Land Change Science*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Hay, G. J. & Castilla, G. (2006). Object-based image analysis: Strengths, weaknesses, opportunities and threats (SWOT), *Proc. 1st Int. Conf. Object-based Image Analysis* (OBIA), 2006. 17.04.2011, Available from: www.commission4.isprs.org/obia06/Papers/01_Opening%20Session/OBIA2006_Hay_Castilla.pdf

Hudelot, C.; Atif, J. & Bloch, I. (2008). Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets and Systems Archive*, Vol. 159 , No. 15, pp. 1929-1951.

Jain, A. K.; Duin, R. & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Trans. Pattern. Anal. Machine Intell.*, Vol. 22, No. 1, pp. 4-37.

Jain, A. & Healey, G. (1998). A multiscale representation including opponent color features for texture recognition. *IEEE Trans. Image Proc.*, Vol. 7, No. 1, pp. 124-128.

Jordan, M. & Jacobs, R. (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, Vol. 6, pp.181–214.

Kandel, E. R. (1991). Perception of motion, depth and form, In: *Principles of Neural Science*, E. Kandel & J. Schwartz, (Eds.), pp. 441-466, Appleton and Lange, Norwalk, Connecticut.

Lawley, J. (2003). Self-organising complex-adaptive systems, In: *Large Group Metaphor Process* (at the Findhorn Community), 17.04.2011, Available from: http://www.cleanlanguage.co.uk/articles/articles/216/1/Self-Organising-Systems-Findhorn/Page1.html

Legg, S. & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. 17.04.2011, Available from: http://arxiv.org/abs/0706.3639.Marr, D. (1982). *Vision*. Freeman and C., New York.

Lindeberg, T. (1993). Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *Int. J. Computer Vision*, Vol. 11, No. 3, pp. 283-318.

Lunetta, E. & Elvidge, C. (1999). *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*, Taylor and Francis, London, United Kingdom.

Martinetz, T. & Schulten, K. (1994). Topology representing networks. *Neural Networks*, Vol. 7, No. 3, pp. 507–522.

Mason, C. & Kandel, E. R.. (1991). Central visual pathways, In: *Principles of Neural Science*, E. Kandel & J. Schwartz, (Eds.), pp. 420-439, Appleton and Lange, Norwalk, Connecticut.

Mather, P. (1994). *Computer Processing of Remotely-Sensed Images – An Introduction*. John Wiley & Sons, Chichester, GB.

Matsuyama, T. & Shang-Shouq Hwang, V. (1990). *SIGMA – A Knowledge-based Aerial Image Understanding System*. Plenum Press, New York.

Mitchell, T. (1997). *Machine Learning*, McGraw-Hill, New York.

Page-Jones, M. (1988). *The Practical Guide to Structured Systems Design*, Prentice-Hall, Englewood Cliffs, New Jersey.

Pakzad, K.; Bückner, J. & Growe, S. (1999). Knowledge Based Moorland Interpretation using a Hybrid System for Image Analysis, *Proc. International Society for Photogrammetry and Remote Sensing* (ISPRS) Conf., Munich, Germany, Sept. 8-10, 1999, 17.04.2011, Available from: http://www.tnt.uni-hannover.de/papers/view.php?ind=1999&ord=Authors&mod=ASC

Pekkarinen, A.; Reithmaier, L. & Strobl, P. (2009). Pan-european forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data. *ISPRS J. Photogrammetry & Remote Sens*., Vol. 64, No. 2 (March 2009), pp. 171-183.

Perona, P. & Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. 12, No. 7, pp. 629-639.

Persello, C. & Bruzzone, L. (2010). A novel protocol for accuracy assessment in classification of very high resolution imagery. *IEEE Trans. Geosci. Remote Sensing*, Vol. 48, No. 3, pp. 1232-1244.

Petrou, M. & Sevilla, P. (2006). *Image processing: Dealing with Texture*. John Wiley & Sons, Chichester, Great Britain.

Qiyao Yu & Clausi, D. A. (2007). SAR sea-ice image analysis based on iterative region growing using semantics. *IEEE Trans. Geosci. Remote Sensing*, Vol. 45, no. 12, pp. 3919-3931.

Richter, R. (2006). Atmospheric / Topographic correction for satellite imagery – ATCOR-2/3 User Guide, Version 6.2. 17.04.2011, Available from: http://www.geog.umontreal.ca/donnees/geo6333/atcor23_manual.pdf

R. Laurini and D. Thompson, Fundamentals of Spatial Information Systems, Academic Press, London, 1992.

Sart, F.; Inglada, J.; Landry, R., & Pultz, T (2001).Risk management using remote sensing data: Moving from scientific to operational applications, *Proc. SBSR Workshop*, Brasil, Apr. 23-27, 2001, 17.04.2011, Available from: www.treemail.nl/download/sarti01.pdf

Shunlin Liang. (2004). *Quantitative Remote Sensing of Land Surfaces*. J. Wiley and Sons, Hoboken, New Jersey.

Sjahputera, O.; Davis, C.H.; Claywell, B.; Hudson, N.J.; Keller, J.M.; Vincent, M.G.; Li, Y.; Klaric, M. & Shyu, C.R. (2008). GeoCDX: An automated change detection and exploitation system for high resolution satellite imagery, *Proc. Int. Geoscience and Remote Sensing Symposium* (IGARSS), Boston (MT), July 6-11, 2008, paper no. FR4.101.1.

T.F. Cootes and C.J.Taylor, Imaging Science and Biomedical Engineering, Draft Report, University of Manchester, Manchester, March 8, 2004, Available from: http://www.isbe.man.ac.uk/~bim/Models/app_models.pdf.

Tapsall, B.; Milenov, P. & Tasdemir, K. (2010). Analysis of RapidEye imagery for annual land cover mapping as an aid to European Union (EU) Common agricultural policy, W. Wagner, B. Székely, (Eds.*): ISPRS TC VII Symposium – 100 Years ISPRS*, Vienna, Austria, July 5–7, 2010, IAPRS, Vol. XXXVIII, Part 7B. 17.04.2011, Available from: http://mars.jrc.it/mars/content/download/1648/8982/file/P4-5_Milenov _Tapsall_BG_RapidEye_Project_JRC_ver2.pdf

USGS & NASA. (2011). Web-enabled Landsat data (WELD) Project. 17.04.2011, Available from: http://landsat.usgs.gov/WELD.php

Vecera, S. P. & Farah, M. J. (1997). Is visual image segmentation a bottom-up or an interactive process?. *Perception & Psychophysics*, Vol. 59, pp. 1280–1296.

Wald, L.; Ranchin, T. & Mangolini, M. (1997). Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogramm. Eng. Remote Sens.*, Vol. 63, No. 6, pp. 691-699.

Wang, Z. & Bovik, A. C. (2002). A universal image quality index. *IEEE Signal Proc. Letters*, Vol. 9, No. 3, pp. 81-84.

Wilson, H. R. & Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Res.*, Vol. 19, pp. 19-32.

Wilson, D. R. & Martinez, T. R. (2000). The Inefficiency of batch training for large training sets. *IEEE-INNS-ENNS International Joint Conference on Neural Networks* (IJCNN'00), 2000, vol. 2, pp.2113.Xu, R. & Wunsch II, D. (2005). Survey of clustering algorithms. *IEEE Trans. Neural Netw.*, Vol. 16, No. 3, pp. 645-678.

Yang, J. & Wang, R. S. (2007). Classified road detection from satellite images based on perceptual organization. *Int. J. Remote Sensing*, Vol. 28, No. 20, pp. 4653-4669.

Zamperoni, P. (1996). Plus ça va, moins ça va. *Pattern Recognition Letters*, Vol. 17, No. 7, (1996), pp. 671-677.

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, Vol. 58, No. 4, pp. 479-493.